QikProp 2.5

User Manual

Schrödinger Press

Copyright © 2006 Schrödinger, LLC. All rights reserved. CombiGlide, Epik, Glide, Impact, Jaguar, Liaison, LigPrep, Maestro, Phase, Prime, QikProp, QikFit, QikSim, QSite, SiteMap, and Strike are trademarks of Schrödinger, LLC.
Schrödinger and MacroModel are registered trademarks of Schrödinger, LLC.

Python is a copyrighted work of the Python Software Foundation. All rights reserved.

To the maximum extent permitted by applicable law, this publication is provided "as is" without warranty of any kind. This publication may contain trademarks of other companies.

Please note that any third party programs ("Third Party Programs") or third party Web sites ("Linked Sites") referred to in this document may be subject to third party license agreements and fees. Schrödinger, LLC and its affiliates have no responsibility or liability, directly or indirectly, for the Third Party Programs or for the Linked Sites or for any damage or loss alleged to be caused by or in connection with use of or reliance thereon. Any warranties that we make regarding our own products and services do not apply to the Third Party Programs or Linked Sites, or to the interaction between, or interoperability of, our products and services and the Third Party Programs. Referrals and links to Third Party Programs and Linked Sites do not constitute an endorsement of such Third Party Programs or Linked Sites.

Revision A, April 2006

Contents

Document Conventions

In addition to the use of italics for names of documents, the font conventions that are used in this document are summarized in the table below.

Table 3.1.

Font	Example	Use
Sans serif	Project Table	Names of GUI features, such as panels, menus, menu items, buttons, and labels
Monospace	`$SCHRODINGER/maestro`	File names, directory names, commands, environment variables, and screen output
Italic	*filename*	Text that the user must replace with a value
Sans serif uppercase	CTRL+H	Keyboard keys

In descriptions of command syntax, the following UNIX conventions are used: braces `{ }` enclose a choice of required items, square brackets `[ ]` enclose optional items, and the bar symbol | separates items in a list from which one item must be chosen. Lines of command syntax that wrap should be interpreted as a single command.

In this document, to *type* text means to type the required text in the specified location, and to *enter* text means to type the required text, then press the ENTER key.

References to literature sources are given in square brackets, like this: [10].

The QikProp Product Suite

1.1 QikProp

QikProp™ is a quick, accurate, easy-to-use absorption, distribution, metabolism, and excretion (ADME) prediction program designed by Professor William L. Jorgensen. QikProp predicts physically significant descriptors and pharmaceutically relevant properties of organic molecules, either individually or in batches.

In addition to predicting molecular properties, QikProp provides ranges for comparing a particular molecule's properties with those of 95% of known drugs. QikProp also flags 30 types of reactive functional groups that may cause false positives in high-throughput screening (HTS) assays. The range of values that cause a molecule to be flagged as dissimilar to other known drugs can be modified in the `QPlimits` file—see Section 4.6 on page 52 for details.

QikProp has two processing modes. By default, QikProp runs in *normal* mode, predicting 44 properties for as many as 10,000 molecules in an hour. Alternatively, QikProp can run in *fast* mode, which predicts 40 properties rather than the default 44. In fast mode, QikProp is able to evaluate approximately 300,000 compounds per hour on a 2 GHz Pentium processor. Fast mode skips the PM3 calculation, which produces the dipole moment, ionization potential and electron affinity. As a result, the octanol/gas partition coefficient is different. Fast mode also bypasses some of the tests performed in normal mode, such as checking for a single molecule, neutral functional groups and valid atom types. The properties generated in fast mode for molecules with these violations are not valid, though the descriptors are still valid.

QikProp output can be used as input for the QikFit™ and QikSim™ modules. The QikFit module performs linear regression analyses for experimentally determined molecular properties. The resulting regression equations can then be integrated back into QikProp and used to predict the experimental property for additional, structurally similar molecules. QikSim performs similarity/diversity analyses on lists of structures based on QikProp descriptors.

QikProp can be run either from the Maestro GUI or from the command line.

QikProp properties and descriptors can be used as input to Strike™, which is a collection of chemically-aware statistical tools for examining correlations within data. It can develop and employ QSAR/QSPR models using partial least squares, principal component analysis, and multiple linear regression; generate univariate and bivariate statistics; and perform similarity/diversity analysis in descriptor and 2D-structure space. Strike can be run from Maestro. For further information about Strike please contact your Schrödinger sales representative.

For each successfully processed molecule, QikProp produces the following descriptors and properties. Those that are not predicted in fast mode are marked with a dagger (†). Those whose values differ between fast and normal mode are marked with a double dagger (‡). Additional information on many of these properties can be found in the *QikProp Technical Notes*.

molecule name	Molecule name taken from the title line in the input structure file. If the title line is blank, the input file name is used.
#stars	Number of property or descriptor values that fall outside the 95% range of similar values for known drugs. Outlying descriptors and predicted properties are denoted with asterisks (*) in the `.qpsa` file. A large number of stars suggests that a molecule is less drug-like than molecules with few stars.
#amine	Number of non-conjugated amine groups.
#amidine	Number of amidine and guanidine groups.
#acid	Number of carboxylic acid groups.
#amide	Number of non-conjugated amide groups.
#rotor	Number of non-trivial (not CX3), non-hindered (not alkene, amide, small ring) rotatable bonds.
#rctvFG	Number of reactive functional groups; the specific groups are listed in the *jobname* `.out` file. The presence of these groups can lead to false positives in HTS assays and to decomposition/reactivity/ toxicity problems *in vivo*.
CNS	Predicted central nervous system activity on a –2 (inactive) to +2 (active) scale.
MW	Molecular weight of the molecule.
dipole†	Computed dipole moment of the molecule.
SASA	Total solvent accessible surface area (SASA) in square angstroms using a probe with a 1.4 Å radius.
FOSA	Hydrophobic component of the SASA (saturated carbon and attached hydrogen).
FISA	Hydrophilic component of the SASA (SASA on N, O, and H on heteroatoms).
PISA	π (carbon and attached hydrogen) component of the SASA.
WPSA	Weakly polar component of the SASA (halogens, P, and S).
volume	Total solvent-accessible volume in cubic angstroms using a probe with a 1.4 Å radius.
donorHB	Estimated number of hydrogen bonds that would be donated by the solute to water molecules in an aqueous solution. Values are averages taken over a number of configurations, so they can be non-integer.

accptHB	Estimated number of hydrogen bonds that would be accepted by the solute from water molecules in an aqueous solution. Values are averages taken over a number of configurations, so they can be non-integer.
dip^2/V†	Square of the dipole moment divided by the molecular volume. This is the key term in the Kirkwood-Onsager equation for the free energy of solvation of a dipole with volume V.
ACxDN^.5/SA	Index of cohesive interaction in solids. This term represents the relationship $(accptHB(\sqrt{donorHB}))/(SA)$; see *Bioorg. Med. Chem. Lett.* **2000**, *10*, 1155.
glob	Globularity descriptor, $(4\pi r^2)/(SASA)$, where *r* is the radius of a sphere with a volume equal to the molecular volume. Globularity is 1.0 for a spherical molecule.
QPpolrz	Predicted polarizability in cubic angstroms.
QPlogPC16	Predicted hexadecane/gas partition coefficient.
QPlogPoct‡	Predicted octanol/gas partition coefficient.
QPlogPw	Predicted water/gas partition coefficient.
QPlogPo/w	Predicted octanol/water partition coefficient.
QPlogS	Predicted aqueous solubility, log S. S in moles/liter is the concentration of the solute in a saturated solution that is in equilibrium with the crystalline solid.
CIQPlogS	Conformation-independent predicted aqueous solubility, log S. S in moles/liter is the concentration of the solute in a saturated solution that is in equilibrium with the crystalline solid.
QPlogHERG	Predicted IC_{50} value for blockage of HERG K^+ channels.
QPPCaco	Predicted apparent Caco-2 cell permeability in nm/sec. Caco-2 cells are a model for the gut-blood barrier. Note: QikProp predictions are for non-active transport.
QPlogBB	Predicted brain/blood partition coefficient. Note: QikProp predictions are for orally delivered drugs so, for example, dopamine and serotonin are CNS negative because they are too polar to cross the blood-brain barrier
QPPMDCK	Predicted apparent MDCK cell permeability in nm/sec. MDCK cells are considered to be a good mimic for the blood-brain barrier. Note: QikProp predictions are for non-active transport.
QPlogKp	Predicted skin permeability, log K_p.
IP(ev)†	PM3 calculated ionization potential.
EA(eV)†	PM3 calculated electron affinity.
#metabol‡	The number of likely metabolic reactions.
QPLogKhsa	Prediction of binding to human serum albumin.
HumanOralAbsorption	Predicted qualitative human oral absorption: 1, 2, or 3 for low, medium, or high. The text version is reported in the output.

PercentHuman-OralAbsorption	Predicted human oral absorption on 0 to 100% scale.
SAFluorine	Solvent-accessible surface area of fluorine atoms.
SAamideO	Solvent-accessible surface area of amide oxygen atoms.
PSA	Van der Waals surface area of polar nitrogen and oxygen atoms.
#NandO	Number of nitrogen and oxygen atoms.
RuleOfFive	Number of violations of Lipinski's rule of five.
Jm	Predicted maximum transdermal transport rate, $K_p \times MW \times S$ ($\mu g\ cm^{-2}\ hr^{-1}$). K_p and S are obtained from the aqueous solubility and skin permeability, QPlogKp and QPlogS. This property is only written to the output file; it is not used in any other calculations.

1.2 QikFit

QikFit generates optimized regression equations for the correlation of experimental data with sets of descriptors, such as those produced by QikProp. QikFit output can then be used by QikProp to predict the experimental property for any number of additional, structurally similar molecules. Alternatively, QikFit can be used independently from QikProp.

QikFit can also perform trend vector analyses. Trend vector analysis is generally used when there are a large number of descriptors and the ratio of molecules to descriptors is not high. For regression analyses, this ratio should be at least five.

There is no Maestro interface for QikFit. This module must be run from the command line.

1.3 QikSim

QikSim performs similarity/diversity analyses on structures in the input file based on Euclidean distances and Tanimoto coefficients. All included structures are ranked in comparison to a user-specified probe molecule. Output from QikProp can be used as input for QikSim.

There is no Maestro interface for QikSim. This module must be run from the command line.

1.4 Citing QikProp, QikFit, or QikSim in Publications

The use of this product should be acknowledged in publications as:

QikProp, version 2.5, Schrödinger, LLC, New York, NY, 2005.
QikFit, version 2.5, Schrödinger, LLC, New York, NY, 2005.
QikSim, version 2.5, Schrödinger, LLC, New York, NY, 2005.

Introduction to Maestro

Maestro is the graphical user interface for all of Schrödinger's products: CombiGlide™, Epik™, Glide™, Impact™, Jaguar™, Liaison™, LigPrep™, MacroModel®, Phase™, Prime™, QikProp™, QSite™, and Strike™. It contains tools for building, displaying, and manipulating chemical structures; for organizing, loading, and storing these structures and associated data; and for setting up, monitoring, and visualizing the results of calculations on these structures. This chapter provides a brief introduction to Maestro and some of its capabilities. For more information on any of the topics in this chapter, see the *Maestro User Manual*.

2.1 General Interface Behavior

Most Maestro panels are amodal: more than one panel can be open at a time, and a panel need not be closed for an action to be carried out. Each Maestro panel has a Close button so you can hide the panel from view.

Maestro supports the mouse functions common to many graphical user interfaces. The left button is used for choosing menu items, clicking buttons, and selecting objects by clicking or dragging. This button is also used for resizing and moving panels. The right button displays a shortcut menu. Other common mouse functions are supported, such as using the mouse in combination with the SHIFT or CTRL keys to select a range of items and select or deselect a single item without affecting other items.

In addition, the mouse buttons are used for special functions described later in this chapter. These functions assume that you have a three-button mouse. If you have a two-button mouse, ensure that it is configured for three-button mouse simulation (the middle mouse button is simulated by pressing or holding down both buttons simultaneously).

2.2 Starting Maestro

Before starting Maestro, you must first set the `SCHRODINGER` environment variable to point to the installation directory. To set this variable, enter the following command at a shell prompt:

csh/tcsh: `setenv SCHRODINGER` *installation-directory*

bash/ksh: `export SCHRODINGER=`*installation-directory*

You might also need to set the DISPLAY environment variable, if it is not set automatically when you log in. To determine if you need to set this variable, enter the command:

```
echo $DISPLAY
```

If the response is a blank line, set the variable by entering the following command:

csh/tcsh: `setenv DISPLAY` *display-machine-name*`:0.0`

bash/ksh: `export DISPLAY=`*display-machine-name*`:0.0`

After you set the SCHRODINGER and DISPLAY environment variables, you can start Maestro using the command:

```
$SCHRODINGER/maestro options
```

If you add the $SCHRODINGER directory to your path, you only need to enter the command maestro. Options for this command are given in Section 2.1 of the *Maestro User Manual*.

The directory from which you started Maestro is Maestro's current working directory, and all data files are written to and read from this directory unless otherwise specified (see Section 2.8 on page 27). You can change directories by entering the following command in the command input area (see page 8) of the main window:

```
cd directory-name
```

where *directory-name* is either a full path or a relative path.

2.3 The Maestro Main Window

The Maestro main window is shown in Figure 2.1 on page 7. The main window components are listed below.

The following components are always visible:

- **Title bar**—displays the Maestro version, the project name (if there is one) and the current working directory.
- **Auto-Help**—automatically displays context-sensitive help.
- **Menu bar**—provides access to panels.
- **Workspace**—displays molecular structures and other 3D graphical objects.

The following components can be displayed or hidden by choosing the component from the Display menu. Your choice of which main window components are displayed is persistent between Maestro sessions.

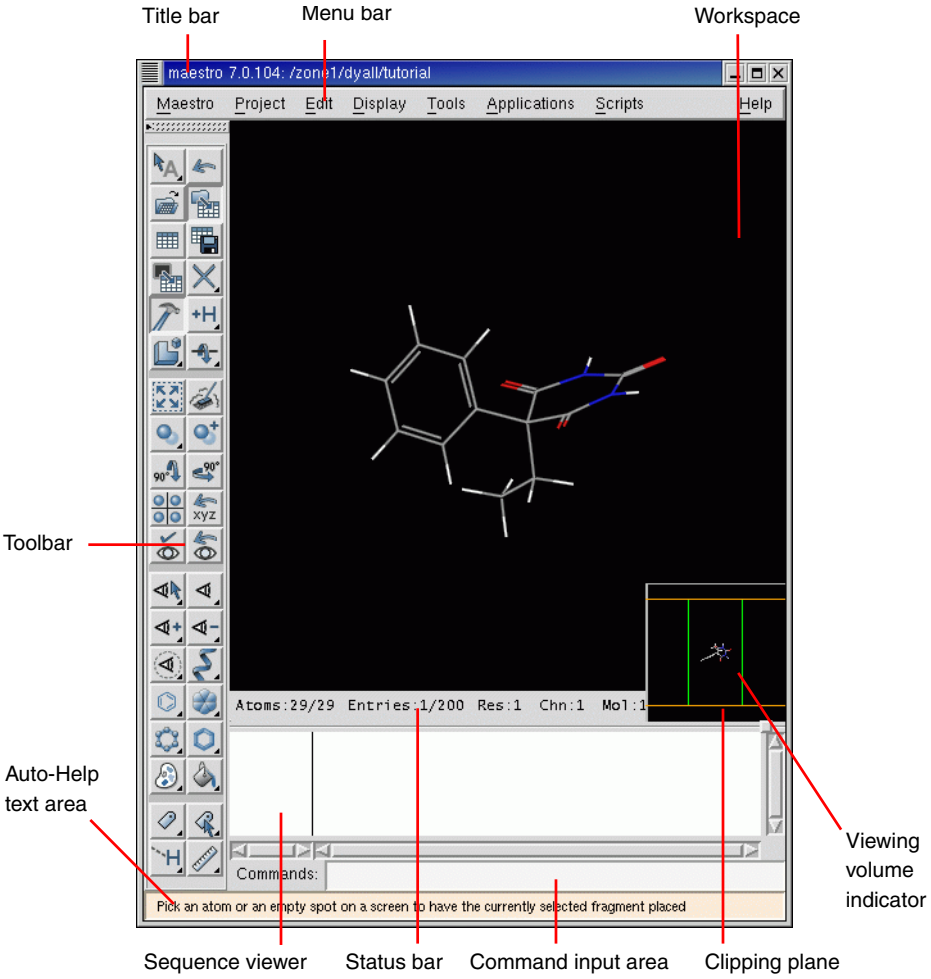

Figure 2.1. The Maestro main window.

- **Toolbar**—contains buttons for many common tasks and provides tools for displaying and manipulating structures, as well as organizing the Workspace.
- **Status bar**—displays information about a particular atom, or about structures in the Workspace, depending on where the pointer pauses (see Section 2.5 of the *Maestro User Manual* for details):
  - **Atom**—displays the chain, residue number, element, PDB atom name, formal charge, and title or entry name (this last field is set by choosing Preferences from the Maestro menu and selecting the Feedback folder).

  - **Workspace**—displays the number of atoms, entries, residues, chains, and molecules in the Workspace.

- **Clipping planes window**—displays a small, top view of the Workspace and shows the clipping planes and viewing volume indicators.
- **Sequence viewer**—shows the sequences for proteins displayed in the Workspace. See Section 2.6 of the *Maestro User Manual* for details.
- **Command input area**—provides a place to enter Maestro commands.

When a distinction between components in the main window and those in other panels is needed, the term *main* is applied to the main window components (e.g., main toolbar).

You can expand the Workspace to occupy the full screen, by pressing CTRL+=. All other components and panels are hidden. To return to the previous display, press CTRL+= again.

2.3.1 The Menu Bar

The menus on the main menu bar provide access to panels, allow you to execute commands, and control the appearance of the Workspace. The main menus are as follows:

- Maestro—save or print images in the Workspace, execute system commands, save or load a panel layout, set preferences, set up Maestro command aliases, and quit Maestro.
- Project—open and close projects, import and export structures, make a snapshot, and annotate a project. These actions can also be performed from the Project Table panel. For more information, see Section 2.4 on page 13.
- Edit—undo actions, build and modify structures, define command scripts and macros, and find atoms in the Workspace.
- Display—control the display of the contents of the Workspace, arrange panels, and display or hide main window components.
- Tools—group atoms; measure, align, and superimpose structures; and view and visualize data.
- Applications—set up, submit, and monitor jobs for Schrödinger's computational programs. Some products have a submenu from which you can choose the task to be performed.
- Scripts—manage and install Python scripts that come with the distribution and scripts that you create yourself. (See Chapter 13 of the *Maestro User Manual* for details.)
- Help—open the Help panel, the PDF documentation index, or information panels; run a demonstration; and display or hide Balloon Help (tooltips).

2.3.2 The Toolbar

The main toolbar contains three kinds of buttons for performing common tasks:

Action—Perform a simple task, like clearing the Workspace.

Display—Open or close a panel or open a dialog box, such as the Project Table panel.

Menu—Display a *button menu*. These buttons have a triangle in the lower right corner.

There are four types of items on button menus, and all four types can be on the same menu (see Figure 2.2):

- **Action**—Perform an action immediately.
- **Display**—Open a panel or dialog box.
- **Object types for selection**—Choose Atoms, Bonds, Residues, Chains, Molecules, or Entries, then click on an atom in the Workspace to perform the action on all the atoms in that structural unit.

  The object type is marked on the menu with a red diamond and the button is indented to indicate the action to be performed.
- **Other setting**—Set a state, choose an attribute, or choose a parameter and click on atoms in the Workspace to display or change that parameter.

The toolbar buttons are described below. Some descriptions refer to features not described in this chapter. See the *Maestro User Manual* for a fuller description of these features.

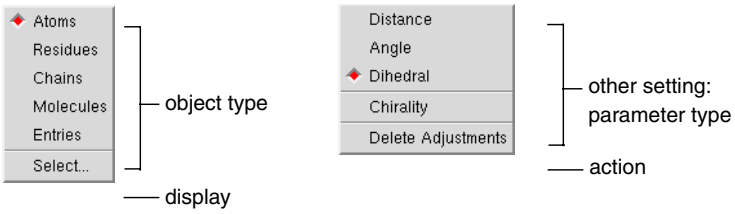

Figure 2.2. The Workspace selection ***button menu and the*** Adjust distances, angles or dihedrals ***button menu.***

Workspace selection
– Choose an object type for selecting
– Open the Atom Selection dialog box

Undo/Redo
Undo or redo the last action. Performs the same function as the Undo item on the Edit menu, and changes to an arrow pointing in the opposite direction when an Undo has been performed, indicating that its next action is Redo.

Open a project
Open the Open Project dialog box.

Import structures
Open the Import panel.

Open/Close Project Table
Open the Project Table panel or close it if it is open.

Save as
Open the Save Project As dialog box, to save the project with a new name.

Create entry from Workspace
Open a dialog box in which you can create an entry in the current project using the contents of the Workspace.

Delete
– Choose an object type for deletion
– Delete hydrogens and waters
– Open the Atom Selection dialog box
– Delete other items associated with the structures in the Workspace
– Click to select atoms to delete
– Double-click to delete all atoms

Open/Close Build panel
Open the Build panel or close it if it is open.

Add hydrogens
– Choose an object type for applying a hydrogen treatment
– Open the Atom Selection dialog box
– Click to select atoms to treat
– Double-click to apply to all atoms

Local transformation
– Choose an object type for transforming
– Click to select atoms to transform
– Open the Advanced Transformations panel

Adjust distances, angles or dihedrals
– Choose a parameter for adjusting
– Delete adjustments

Fit to screen
Scale the displayed structure to fit into the Workspace and reset the center of rotation.

Clear Workspace
Clear all atoms from the Workspace.

Set fog display state
Choose a fog state. Automatic means fog is on when there are more than 40 atoms in the Workspace, otherwise it is off.

Enhance depth cues
Optimize fogging and other depth cues based on what is in the Workspace.

Rotate around X axis by 90 degrees
Rotate the Workspace contents around the X axis by 90 degrees.

Rotate around Y axis by 90 degrees
Rotate the Workspace contents around the Y axis by 90 degrees.

Tile entries
Arrange entries in a rectangular grid in the Workspace.

Reset Workspace
Reset the rotation, translation, and zoom of the Workspace to the default state.

Save view
Save the current view of the Workspace: orientation, location, and zoom.

Restore view
Restore the last saved view of the Workspace: orientation, location, and zoom.

Display only selected atoms
– Choose an object type for displaying
– Click to select atoms to display
– Double-click to display all atoms

Display only
– Choose a predefined atom category
– Open the Atom Selection dialog box

Also display
– Choose a predefined atom category
– Open the Atom Selection dialog box

Undisplay
– Choose a predefined atom category
– Open the Atom Selection dialog box

Display residues within N angstroms of currently displayed atoms
– Choose a radius
– Open a dialog box to set a value

Show, hide, or color ribbons
– Choose to show or hide ribbons
– Choose a color scheme for coloring ribbons

Draw bonds in wire
– Choose an object type for drawing bonds in wire representation
– Open the Atom Selection dialog box
– Click to select atoms for representation
– Double-click to apply to all atoms

Draw atoms in CPK
– Choose an object type for drawing bonds in CPK representation
– Open the Atom Selection dialog box
– Click to select atoms for representation
– Double-click to apply to all atoms

Draw atoms in Ball & Stick
– Choose an object type for drawing bonds in Ball & Stick representation
– Open the Atom Selection dialog box
– Click to select atoms for representation
– Double-click to apply to all atoms

Draw bonds in tube
– Choose an object type for drawing bonds in tube representation
– Open the Atom Selection dialog box
– Click to select atoms for representation
– Double-click to apply to all atoms

Color all atoms by scheme
Choose a predefined color scheme.

Color residue by constant color
– Choose a color for applying to residues
– Click to select residues to color
– Double-click to color all atoms

Label atoms
– Choose a predefined label type
– Delete labels

Label picked atoms
– Choose an object type for labeling atoms
– Open the Atom Selection dialog box
– Open the Atom Labels panel at the Composition folder
– Delete labels
– Click to select atoms to label
– Double-click to label all atoms

Display H-bonds
– Choose bond type:
intra—displays H-bonds within the selected molecule
inter—displays H-bonds between the selected molecule and all other atoms.
– Delete H-bonds
– Click to select molecule

Measure distances, angles or dihedrals
– Choose a parameter for displaying measurements
– Delete measurements
– Click to select atoms for measurement

2.3.3 Mouse Functions in the Workspace

The left mouse button is used for selecting objects. You can either click on a single atom or bond, or you can drag to select multiple objects. The right mouse button opens shortcut menus, which are described in Section 2.7 of the *Maestro User Manual*.

The middle and right mouse buttons can be used on their own and in combination with the SHIFT and CTRL keys to perform common operations, such as rotating, translating, centering, adjusting, and zooming.

Table 2.1. Mapping of Workspace operations to mouse actions.

Mouse Button	Keyboard	Motion	Action
Left		click, drag	Select
Left	SHIFT	click, drag	Toggle the selection
Middle		drag	Rotate about X and Y axes Adjust bond, angle, or dihedral
Middle	SHIFT	drag vertically	Rotate about X axis
Middle	SHIFT	drag horizontally	Rotate about Y axis
Middle	CTRL	drag horizontally	Rotate about Z axis
Middle	SHIFT + CTRL	drag horizontally	Zoom
Right		click	Spot-center on selection
Right		click and hold	Display shortcut menu
Right		drag	Translate in the X-Y plane
Right	SHIFT	drag vertically	Translate along the X axis
Right	SHIFT	drag horizontally	Translate along the Y axis
Right	CTRL	drag horizontally	Translate along the Z axis
Middle & Right		drag horizontally	Zoom

2.3.4 Shortcut Key Combinations

Some frequently used operations have been assigned shortcut key combinations. The shortcuts available in the main window are described in Table 2.2.

Table 2.2. Shortcut keys in the Maestro main window.

Keys	Action	Equivalent Menu Choices
CTRL+B	Open Build panel	Edit > Build
CTRL+C	Create entry	Project > Create Entry From Workspace
CTRL+E	Open Command Script Editor panel	Edit > Command Script Editor
CTRL+F	Open Find Atoms panel	Edit > Find
CTRL+H	Open Help panel	Help > Help
CTRL+I	Open Import panel	Project > Import Structures
CTRL+M	Open Measurements panel	Tools > Measurements
CTRL+N	Create new project	Project > New
CTRL+O	Open project	Project > Open
CTRL+P	Print	Maestro > Print
CTRL+Q	Quit	Maestro > Quit
CTRL+S	Open Sets panel	Tools > Sets
CTRL+T	Open Project Table panel	Project > Show Table
CTRL+W	Close project	Project > Close
CTRL+Z	Undo/Redo last command	Edit > Undo/Redo
CTRL+=	Enter and exit full screen mode (Workspace occupies full screen)	None

2.4 Maestro Projects

All the work you do in Maestro is done within a *project*. A project consists of a set of *entries*, each of which contains one or more chemical structures and their associated data. In any Maestro session, there can be only one Maestro project open. If you do not specify a project when you start Maestro, a *scratch* project is created. You can work in a scratch project without saving it, but you must save it in order to use it in future sessions. When you save or close a project, all the view transformations (rotation, translation, and zoom) are saved with it. When you close a project, a new scratch project is automatically created.

Likewise, if there is no entry displayed in the Workspace, Maestro creates a *scratch* entry. Structures that you build in the Workspace constitute a scratch entry until you save the structures as project entries. The scratch entry is not saved with the project unless you explicitly add it to the project. However, you can use a scratch entry as input for some calculations.

To add a scratch entry to a project, do one of the following:

- Click the Create entry from Workspace button:

- Choose Create Entry from Workspace from the Project menu.
- Press CTRL+C.

In the dialog box, enter a name and a title for the entry. The entry name is used internally to identify the entry and can be modified by Maestro. The title can be set or changed by the user, but is not otherwise modified by Maestro.

Once an entry has been incorporated into the project, its structures and their data are represented by a row in the Project Table. Each row contains the row number, an icon indicating whether the entry is displayed in the Workspace (the In column), the entry title, a button to open the Surfaces panel if the entry has surfaces, the entry name, and any entry properties. The row number is not a property of the entry.

Entries can be collected into groups, and the members of the group can be displayed or hidden. Most additions of multiple entries to the Project Table are done as entry groups.

You can use entries as input for all of the computational programs—Glide, Impact, Jaguar, Liaison, LigPrep, MacroModel, Phase, Prime, QikProp, QSite, and Strike. You can select entries as input for the ePlayer, which displays the selected structures in sequence. You can also duplicate, combine, rename, and sort entries; create properties; import structures as entries; and export structures and properties from entries in various formats.

To open the Project Table panel, do one of the following:

- Click the Open/Close Project Table button on the toolbar

- Choose Show Table from the Project menu
- Press CTRL+T.

The Project Table panel contains a menu bar, a toolbar, and the table itself.

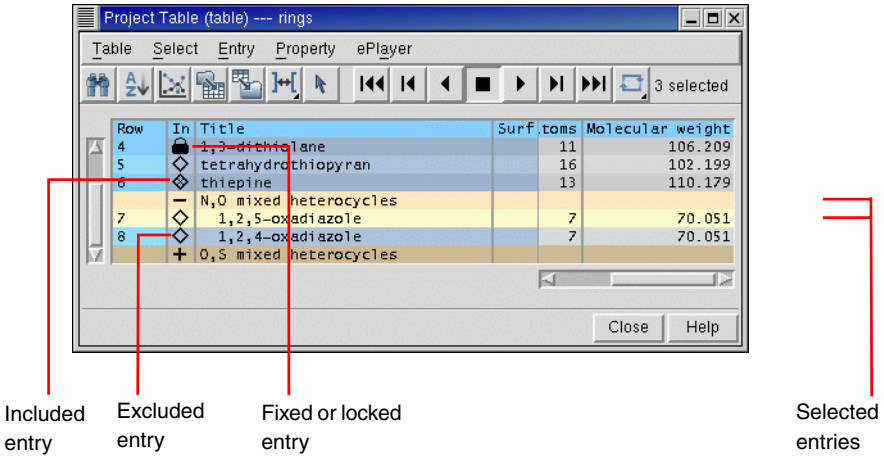

Figure 2.3. The** Project Table **panel.

2.4.1 The Project Table Toolbar

The Project Table toolbar contains two groups of buttons and a status display. The first set of buttons opens various panels that allow you to perform functions on the entries in the Project Table. The second set of buttons controls the ePlayer, which "plays through" the selected structures: each structure is displayed in the Workspace in sequence, at a given time interval. See Section 2.3.2 on page 9 for a description of the types of toolbar buttons. The buttons are described below.

Find
Open the Find panel for locating alphanumeric text in any column of the Project Table, except for the row number.

Sort
Open the Sort panel for sorting entries by up to three properties.

Plot
Open the Plot panel for plotting entry properties.

Import Structure
Open the Import panel for importing structures into the project.

Export Structure
Open the Export panel for exporting structures to a file.

Columns
Choose an option for adjusting the column widths.

Select only
Open the Entry Selection dialog box for selecting entries based on criteria for entry properties.

Go to start
Display the first selected structure.

Previous
Display the previous structure in the list of selected structures.

Play backward
Display the selected structures in sequence, moving toward the first.

Stop
Stop the ePlayer.

Play forward
Display the selected structures in sequence, moving toward the last.

Next
Display the next structure in the list of selected structures.

Go to end
Display the last selected structure.

Loop
Choose an option for repeating the display of the structures. Single Direction displays structures in a single direction, then repeats. Oscillate reverses direction each time the beginning or end of the list is reached.

The status display, to the right of the toolbar buttons, shows the number of selected entries. When you pause the cursor over the status display, the Balloon Help shows the total number of entries, the number shown in the table, the number selected, and the number included in the Workspace.

2.4.2 The Project Table Menus

- Table—find text, sort entries, plot properties, import and export structures, and configure the Project Table.
- Select—select all entries, none, invert your selection, or select classes of entries using the Entry Selection dialog box and the Filter panel.

- Entry—include or exclude entries from the Workspace, display or hide entries in the Project Table, and perform various operations on the selected entries.
- Property—display and manipulate entry properties in the Project Table.
- ePlayer—view entries in succession, stop, reverse, and set the ePlayer options.

2.4.3 Selecting Entries

Many operations in Maestro are performed on the entries selected in the Project Table. The Project Table functions much like any other table: select rows by clicking, shift-clicking, and control-clicking. However, because clicking in an editable cell of a selected row enters edit mode, you should click in the Row column to select entries. See Section 2.4.5 on page 18 for more information on mouse actions in the Project Table. There are shortcuts for selecting classes of entries on the Select menu.

In addition to selecting entries manually, you can select entries that meet a combination of conditions on their properties. Such combinations of conditions are called *filters*. Filters are Entry Selection Language (ESL) expressions and are evaluated at the time they are applied. For example, if you want to set up a Glide job that uses ligands with a low molecular weight (say, less than 300) and that has certain QikProp properties, you can set up a filter and use it to select entries for the job. If you save the filter, you can use it again on a different set of ligands that meet the same selection criteria.

To create a filter:

1. Do one of the following:
   - Choose Only, Add, or Deselect from the Select menu.
   - Click the Entry selection button on the toolbar.

   
2. In the Properties folder, select a property from the property list, then select a condition.
3. Combine this selection with the current filter by clicking Add, Subtract, or Intersect. These buttons perform the Boolean operations OR, AND NOT, and AND on the corresponding ESL expressions.
4. To save the filter for future use click Create Filter, enter a name, and click OK.
5. Click OK to apply the filter immediately.

2.4.4 Including Entries in the Workspace

In addition to selecting entries, you can also use the Project Table to control which entries are displayed in the Workspace. An entry that is displayed in the Workspace is *included* in the Workspace; likewise, an entry that is not displayed is *excluded*. Included entries are marked by an X in the diamond in the In column; excluded entries are marked by an empty diamond. Entry inclusion is completely independent of entry selection.

To include or exclude entries, click, shift-click, or control-click in the In column of the entries, or select entries and choose Include or Exclude from the Entry menu. Inclusion with the mouse works just like selection: when you include an entry by clicking, all other entries are excluded.

It is sometimes useful to keep one entry in the Workspace and include others one by one: for example, a receptor and a set of ligands. You can fix the receptor in the Workspace by selecting it in the Project Table and choosing Fix from the Entry menu or by pressing CTRL+F. A padlock icon replaces the diamond in the In column to denote a *fixed* entry. To remove a fixed entry from the Workspace, you must exclude it explicitly (CTRL+X). It is not affected by the inclusion or exclusion of other entries. Fixing an entry affects only its inclusion; you can still rotate, translate, or modify the structure.

2.4.5 Mouse Functions in the Project Table

The Project Table supports the standard use of shift-click and control-click to select objects. This behavior applies to the selection of entries and the inclusion of entries in the Workspace. You can also drag to resize rows and columns and to move rows.

You can drag a set of non-contiguous entries to reposition them in the Project Table. When you release the mouse button, the entries are placed after the first unselected entry that precedes the entry on which the cursor is resting. For example, if you select entries 2, 4, and 6, and release the mouse button on entry 3, these three entries are placed after entry 1, because entry 1 is the first unselected entry that precedes entry 3. To move entries to the top of the table, drag them above the top of the table; to move entries to the end of the table, drag them below the end of the table.

A summary of mouse functions in the Project Table is provided in Table 2.3.

Table 2.3. Mouse operations in the Project Table.

Task	Mouse Operation
Change a Boolean property value	Click repeatedly in a cell to cycle through the possible values (On, Off, Clear)
Display the Entry menu for an entry	Right-click anywhere in the entry. If the entry is not selected, it becomes the selected entry. If the entry is selected, the action is applied to all selected entries.
Display a version of the Property menu for a property	Right-click in the column header
Edit the text or the value in a table cell	Click in the cell and edit the text or value
Include an entry in the Workspace, exclude all others	Click the In column of the entry
Move selected entries	Drag the entries
Paste text into a table cell	Middle-click
Resize rows or columns	Drag the boundary with the middle mouse button
Select an entry, deselect all others	For an unselected entry, click anywhere in the row except the In column; for a selected entry, click the row number.
Select or include multiple entries	Click the first entry then shift-click the last entry
Toggle the selection or inclusion state	Control-click the entry or the In column

2.4.6 Project Table Shortcut Keys

Some frequently used project operations have been assigned shortcut key combinations. The shortcuts, their functions, and their menu equivalents are listed in Table 2.4.

Table 2.4. Shortcut keys in the Project Table.

Keys	Action	Equivalent Menu Choices
CTRL+A	Select all entries	Select > All
CTRL+F	Fix entry in Workspace	Entry > Fix
CTRL+I	Open Import panel	Table > Import Structures
CTRL+N	Include only selected entries	Entry > Include Only
CTRL+U	Deselect all entries	Select > None
CTRL+X	Exclude selected entries	Entry > Exclude
CTRL+Z	Undo/Redo last command	Edit > Undo/Redo in main window

2.5 Building a Structure

After you start Maestro, the first task is usually to create or import a structure. You can open existing Maestro projects or import structures from other sources to obtain a structure, or you can build your own. To open the Build panel, do one of the following:

- Click the Open/Close Build panel button in the toolbar:

- Choose Build from the Edit menu.
- Press CTRL+B.

The Build panel allows you to create structures by drawing or placing atoms or fragments in the Workspace and connecting them into a larger structure, to adjust atom positions and bond orders, and to change atom properties. This panel contains a toolbar and three folders.

2.5.1 Placing and Connecting Fragments

The Build panel provides several tools for creating structures in the Workspace. You can place and connect fragments, or you can draw a structure freehand.

To place a fragment in the Workspace:

1. Select Place.
2. Choose a fragment library from the Fragments menu.
3. Click a fragment.
4. Click in the Workspace where you want the fragment to be placed.

To connect fragments in the Workspace, do one of the following:

- Place another fragment and connect them using the Connect & Fuse panel, which you open from the Edit menu on the main menu bar or with the Display Connect & Fuse panel on the Build toolbar.

- Replace one or more atoms in the existing fragment with another fragment by selecting a fragment and clicking in the Workspace on the main atom to be replaced.
- Grow another fragment by selecting Grow in the Build panel and clicking the fragment you want to add in the Fragments folder.

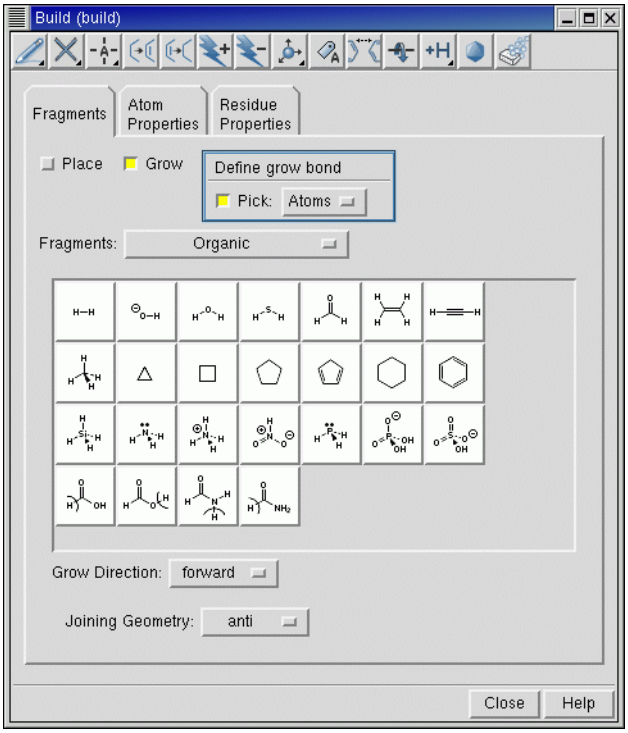

Figure 2.4. The* Build *panel.

Grow mode uses predefined rules to connect a fragment to the *grow bond*. The grow bond is marked by a green arrow. The new fragment replaces the atom at the head of the arrow on the grow bond and all atoms attached to it. To change the grow bond, choose Bonds from the Pick option menu in the Build panel and click on the desired grow bond in the Workspace. The arrow points to the atom nearest to where you clicked.

To draw a structure freehand:

1. Choose an element from the Draw button menu on the Build panel toolbar:

2. Click in the Workspace to place an atom of that element.
3. Click again to place another atom and connect it to the previous atom.
4. Continue this process until you have drawn the structure.
5. Click the active atom again to finish drawing.

2.5.2 Adjusting Properties

In the Atom Properties folder, you can change the properties of the atoms in the Workspace. For each item on the Property option menu—Element, Atom Type (MacroModel), Partial Charge, PDB Atom Name, Grow Name, and Atom Name—there is a set of tools you can use to change the atom properties. For example, the Element tools consist of a periodic table from which you can choose an element and select an atom to change it to an atom of the selected element.

Similarly, the Residue Properties folder provides tools for changing the properties of residues: the Residue Number, the Residue Name, and the Chain Name.

To adjust bond lengths, bond angles, dihedral angles, and chiralities during or after building a structure, use the Adjust distances, angles or dihedrals button on the main toolbar:

You can also open the Adjust panel from this button menu, from the Display Adjust panel button on the Build panel toolbar (which has the same appearance as the above button) or from the Edit menu in the main window.

2.5.3 The Build Panel Toolbar

The toolbar of the Build panel provides quick access to tools for drawing and modifying structures and labeling atoms. See Section 2.3.2 on page 9 for a description of the types of toolbar buttons. The toolbar buttons and their use are described below.

Free-hand drawing
Choose an element for drawing structures freehand in the Workspace (default C). Each click in the Workspace places an atom and connects it to the previous atom.

Delete
Choose an object for deleting. Same as the Delete button on the main toolbar, see page 10.

Set element
Choose an element for changing atoms in the Workspace (default C). Click an atom to change it to the selected element.

Increment bond order
Select a bond to increase its bond order by one, to a maximum of 3.

Decrement bond order
Select a bond to decrease its bond order by one, to a minimum of 0.

Increment formal charge
Select an atom to increase its formal charge by one.

Decrement formal charge
Select an atom to decrease its formal charge by one.

Move
Choose a direction for moving atoms, then click the atom to be moved. Moves in the XY plane are made by clicking the new location. Moves in the Z direction are made in 0.5 Å increments.

Label
Apply heteroatom labels as you build a structure. The label consists of the element name and formal charge, and is applied to atoms other than C and H.

Display Connect & Fuse panel
Open the Connect & Fuse panel so you can connect structures (create bonds between structures) or fuse structures (replace atoms of one structure with those of another).

Display Adjust panel
Open the Adjust panel so you can change bond lengths, bond angles, dihedral angles, or atom chiralities.

Add hydrogens
Choose an atom type for applying the current hydrogen treatment. Same as the Add hydrogens button on the main toolbar, see page 10.

Geometry Symmetrizer
Open the Geometry Symmetrizer panel for symmetrizing the geometry of the structure in the Workspace.

Geometry Cleanup
Clean up the geometry of the structure in the Workspace.

2.6 Selecting Atoms

Maestro has a powerful set of tools for selecting atoms in a structure: toolbar buttons, picking tools in panels, and the Atom Selection dialog box. These tools allow you to select atoms in two ways:

- Select atoms first and apply an action to them
- Choose an action first and then select atoms for that action

2.6.1 Toolbar Buttons

The small triangle in the lower right corner of a toolbar button indicates that the button contains a menu. Many of these buttons allow you to choose an object type for selecting: choose Atoms, Bonds, Residues, Chains, Molecules, or Entries, then click on an atom in the Workspace to perform the action on all the atoms in that structural unit.

For example, to select atoms with the Workspace selection toolbar button:

1. Choose Residues from the Workspace selection button menu:

   

   The button changes to:

   

2. Click on an atom in a residue in the Workspace to select all the atoms in that residue.

2.6.2 Picking Tools

The picking tools are embedded in each panel in which you need to select atoms to apply an operation. The picking tools in a panel can include one or more of the following:

- Pick option menu—Allows you to choose an object type. Depending on the operation to be performed, you can choose Atoms, Bonds, Residues, Chains, Molecules, or Entries, then click on an atom in the Workspace to perform the action on all the atoms in that structural unit.

  The Pick option menu varies from panel to panel, because not all object types are appropriate for a given operation. For example, some panels have only Atoms and Bonds in the Pick option menu.

- All button—Performs the action on all atoms in the Workspace.
- Selection button—Performs the action on any atoms already selected in the Workspace.
- Previous button—Performs the action on the most recent atom selection defined in the Atom Selection dialog box.
- Select button—Opens the Atom Selection dialog box.
- ASL text box—Allows you to type in an ASL expression for selecting atoms.

  ASL stands for Atom Specification Language, and is described in detail in the *Maestro Command Reference Manual*.

- Clear button—Clears the current selection

  

- Show markers option—Marks the selected atoms in the Workspace.

For example, to label atoms with the Label Atoms panel:

1. Choose Atom Labels from the Display menu.
2. In the Composition folder, select Element and Atom Number.
3. In the picking tools section at the top of the panel, you could do one of the following:
   - Click Selection to apply labels to the atoms already selected in the Workspace (from the previous example).
   - Choose Residues from the Pick option menu and click on an atom in a different residue to label all the atoms in that residue.

2.6.3 The Atom Selection Dialog Box

If you wish to select atoms based on more complex criteria, you can use the Atom Selection dialog box. To open this dialog box, choose Select from a button menu or click the Select button in a panel. See Section 5.3 of the *Maestro User Manual* for detailed instructions on how to use the Atom Selection dialog box.

2.7 Scripting in Maestro

Although you can perform nearly all Maestro-supported operations through menus and panels, you can also perform operations using Maestro commands, or compilations of these commands, called *scripts*. Scripts can be used to automate lengthy procedures or repetitive tasks and can be created in several ways. These are summarized below.

2.7.1 Python Scripts

Python is a full-featured scripting language that has been embedded in Maestro to extend its scripting facilities. The Python capabilities within Maestro include access to Maestro functionality for dealing with chemical structures, projects, and Maestro files.

The two main Python commands used in Maestro are:

- `pythonrun`—executes a Python module. (You can also use the alias `pyrun`.) The syntax is:

  `pythonrun` *module*`.`*function*

- `pythonimport`—rereads a Python file so that the next time you use the `pythonrun` command, it uses the updated version of the module. (You can also use the alias `pyimp`.)

From the Maestro Scripts menu you can install, manage, and run Python scripts. For more information on the Scripts menu, see Section 13.1 of the *Maestro User Manual*.

For more information on using Python with Maestro, see *Maestro Scripting with Python*.

2.7.2 Command Scripts

All Maestro commands are logged and displayed in the Command Script Editor panel. This means you can create a command script by performing the operations with the GUI controls, copying the logged commands from the Command History list into the Script text area of the panel, then saving the list of copied commands as a script.

To run an existing command script:

1. Open the Command Script Editor panel from the Edit menu in the main window.
2. Click Open Local and navigate to the directory containing the desired script.
3. Select a script in the Files list and click Open.

   The script is loaded into the Script window of the Command Script Editor panel.
4. Click Run Script.

Command scripts cannot be used for Prime operations.

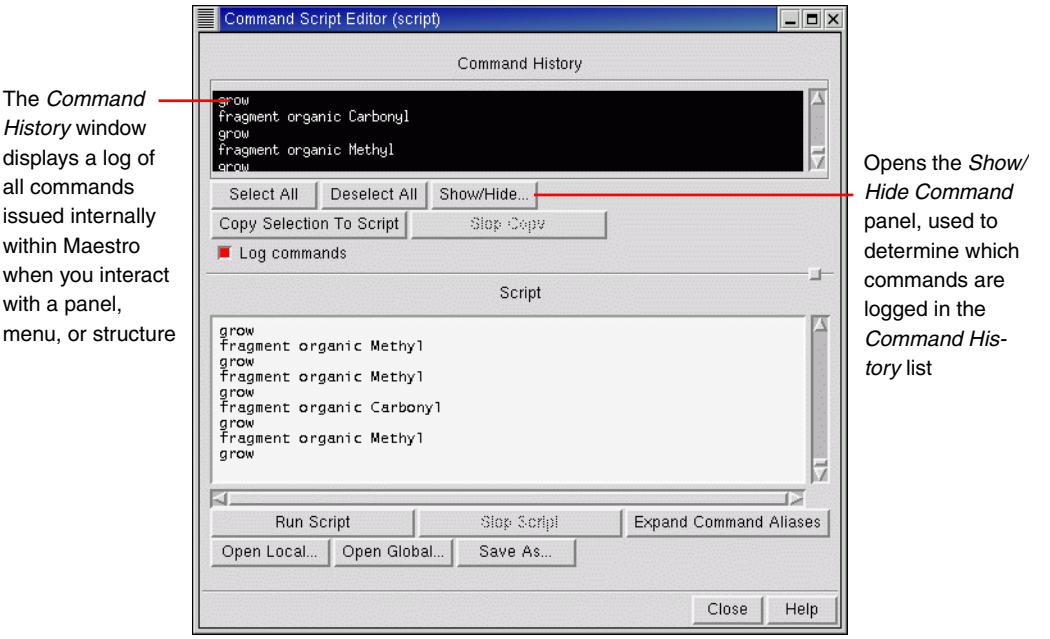

Figure 2.5. The Command Script Editor ***panel.***

2.7.3 Macros

There are two kinds of macros you can create: named macros and macros assigned to function keys F1 through F12.

To create and run a named macro:

1. Open the Macros panel from the Edit menu in the main window.
2. Click New, enter a name for the macro, and click OK.
3. In the Definition text box, type the commands for the macro.
4. Click Update to update the macro definition.
5. To run the macro, enter the following in the command input area in the main window:

   `macrorun` *macro-name*

   If the command input area is not visible, choose Command Input Area from the Display menu.

To create and run a function key macro:

1. Open the Function Key Macros panel from the Edit menu in the main window.
2. From the Macro Key option, select a function key (F1 through F12) to which to assign the macro.
3. In the text box, type the commands for the macro.
4. Click Run to test the macro or click Save to save it.
5. To run the macro from the main window, press the assigned function key.

For more information on macros, see Section 13.5 of the *Maestro User Manual*.

2.8 Specifying a Maestro Working Directory

When you use Maestro to launch QikProp jobs, Maestro writes job output to the directory specified in the Directory folder of the Preferences panel. By default, this directory (the file I/O directory) is the directory from which you started Maestro.

To change the Maestro working directory:

1. Open the Preferences panel from the Maestro menu.
2. Click the Directory tab.
3. Select the directory you want to use for reading and writing files.

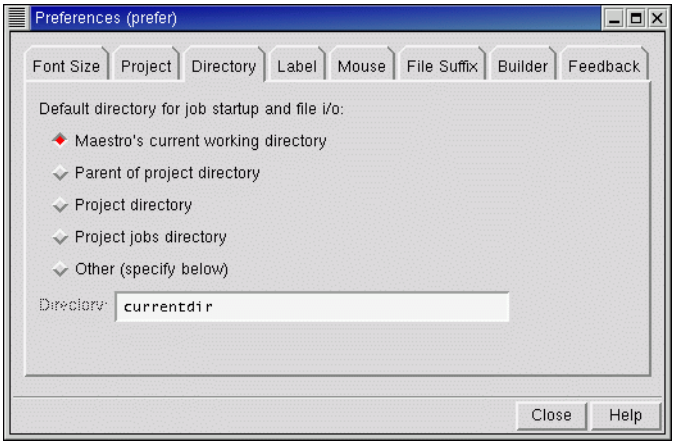

Figure 2.6. T

You can also set other preferences in the Preferences panel. See Section 12.2 of the *Maestro User Manual* for details.

2.9 Undoing an Operation

To undo a single operation, click the Undo button in the toolbar, choose Undo from the Edit menu, or press CTRL+Z. The word Undo in the menu is followed by text that describes the operation to undo. Not all operations can be undone: for example, global rotations and translations are not undoable operations. For such operations you can use the Save view and Restore view buttons in the toolbar, which save and restore a molecular orientation.

2.10 Running and Monitoring Jobs

Maestro has panels for each product for preparing and submitting jobs. To use these panels, choose the appropriate product and task from the Applications menu and its submenus. Set the appropriate options in the panel, then click Start to open the Start dialog box and set options for running the job. For a complete description of the Start dialog box associated with your computational program, see your product's User Manual. When you have finished setting the options, click Start to launch the job and open the Monitor panel.

The Monitor panel is the control panel for monitoring the progress of jobs and for pausing, resuming, or killing jobs. All jobs that belong to your user ID can be displayed in the Monitor panel, whether or not they were started from Maestro. Subjobs are indented under their parent in the job list. The text pane shows various output information from the monitored job, such as the contents of the log file. The Monitor panel opens automatically when you start a job. If it is

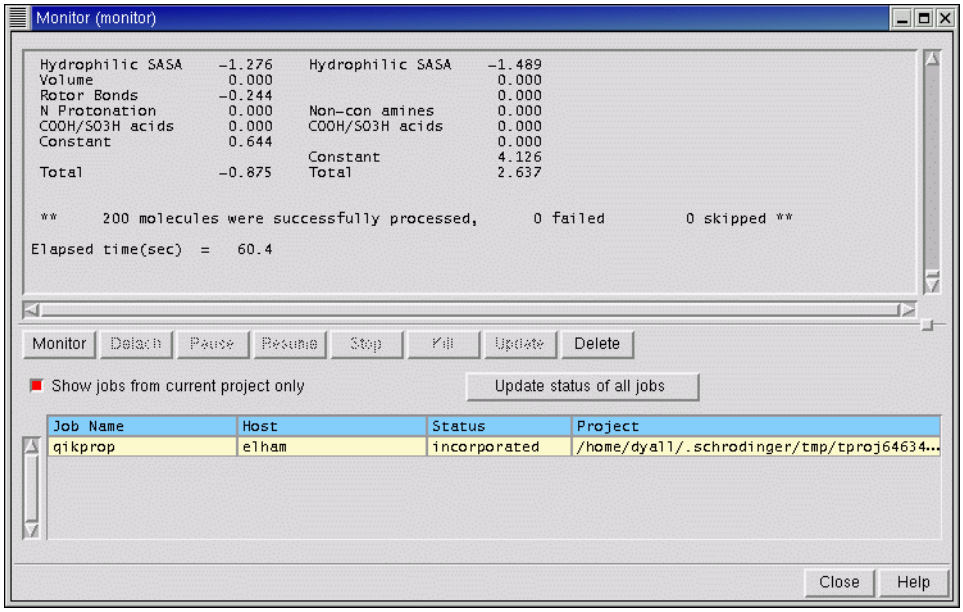

Figure 2.7. The* Monitor *panel.

not open, you can open it by choosing Monitor from the Applications menu in the Maestro main window.

While jobs are running, the Detach, Pause, Resume, Stop, Kill, and Update buttons are active. When there are no jobs currently running, only the Monitor and Delete buttons are active. These buttons act on the selected job. By default, only jobs started from the current project are shown. To show other jobs, deselect Show jobs from current project only.

When a monitored job ends, the results are incorporated into the project according to the settings used to launch the job. If a job that is not currently being monitored ends, you can select it in the Monitor panel and click Monitor to incorporate the results. Monitored jobs are incorporated only if they are part of the current project. You can monitor jobs that are not part of the current project, but their results are not incorporated. To add their results to a project, you must open the project and import the results.

Further information on job control, including configuring your site, monitoring jobs, running jobs, and job incorporation, can be found in the *Job Control Guide* and the *Installation Guide*.

2.11 Getting Help

Maestro comes with automatic, context-sensitive help (Auto-Help), Balloon Help (tooltips), an online help facility, and a user manual. To get help, follow the steps below:

- Check the Auto-Help text box at the bottom of the main window. If help is available for the task you are performing, it is automatically displayed there. It describes what actions are needed to perform the task.
- If your question concerns a GUI element, such as a button or option, there may be Balloon Help for the item. Pause the cursor over the element. If the Balloon Help does not appear, check that Show Balloon Help is selected in the Help menu of the main window. If there is Balloon Help for the element, it appears within a few seconds.
- If you do not find the help you need using either of the steps above, click the Help button in the lower right corner of the appropriate panel. The Help panel is displayed with a relevant help topic.
- For help with a concept or action not associated with a panel, open the Help panel from the Help menu or press CTRL+H.

If you do not find the information you need in the Maestro help system, check the following sources:

- The *Maestro User Manual*
- The Frequently Asked Questions page, found at http://www.schrodinger.com/Support/faq.html

You can also contact Schrödinger by e-mail or phone for help:

- E-mail: help@schrodinger.com
- Phone: (503) 299-1150

2.12 Ending a Maestro Session

To end a Maestro session, choose Quit from the Maestro menu. To save a log file with a record of all operations performed in the current session, click Quit, save log file in the Quit panel. This information can be useful to Schrödinger support staff when responding to any problem you report.

QikProp Tutorial

This chapter is designed to assist you to rapidly become familiar with the capabilities of QikProp 2.5. Once you have worked through these exercises, you will have an understanding of the basic QikProp, QikFit, and QikSim features.

This chapter assumes that you have already installed the Schrödinger software, or that you have access to an installation of the Schrödinger software. If you do not have access to the software, you must install it. For installation instructions, see the *Installation Guide*. It is also assumed that you have set the `SCHRODINGER` environment variable according to the instructions in Section 2.2 on page 5.

3.1 Making Preparations

Before you can run QikProp from Maestro, you must first launch a Maestro session. To do so, use the instructions in Section 2.2 on page 5.

Specifying a Maestro Working Directory

When you use Maestro to launch QikProp jobs, Maestro writes job output to the directory specified in the Directory folder of the Preferences panel. By default, the file i/o directory (the directory to which Maestro will write files) is the directory from which you launched the Maestro program. If you want output files to be placed in another directory, you can change the preferences as described in Section 2.8 on page 27.

Creating a Working Tutorial Directory

In the `$SCHRODINGER/qikprop-v`*version*`/tutorial` directory of your QikProp distribution are some files that you will use in the following exercises. Before you begin the tutorial, you must create a working directory in which to write the files that are produced during the exercises. To create this directory, do the following:

1. Move to a directory in which you have write permission.
2. Create a working directory:

   `mkdir` *workdir*

3. Copy the files to your working directory:

   `cp $SCHRODINGER/qikprop-v`*version*`/tutorial/*.*` *workdir*

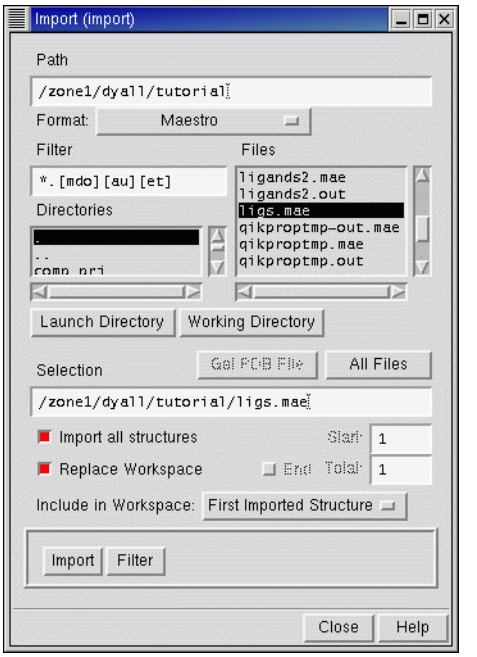

Figure 3.1. The** Import **panel.

3.2 Importing Structures to Use as QikProp Input

When running QikProp from Maestro, you can use as input for a job either the structures in the Workspace or entries that are selected in the Project Table panel. In this exercise, you will import structures into Maestro to use as QikProp input.

1. Click the Import Structures button in the toolbar.

   

   The Import panel is displayed.

2. Select *workdir*`/ligs.mae`.

3. Click Import.

   The structures are imported, and the first structure is displayed in the Workspace.

4. Click the Open/Close Project Table button in the toolbar.

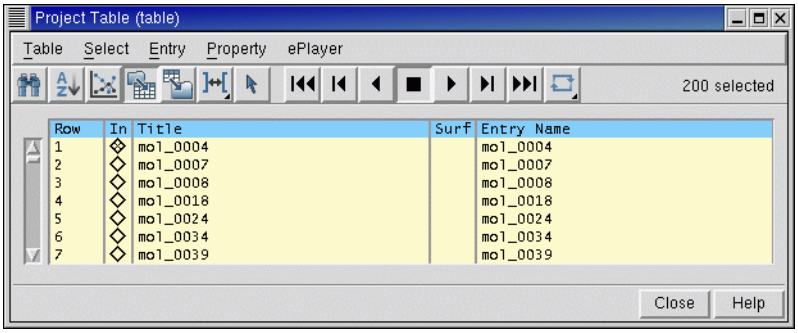

Figure 3.2. The** Project Table **panel with all entries selected.

The Project Table panel is displayed. The contents of the `ligs.mae` file are listed, and all 200 entries are selected.

3.3 Using QikProp to Generate a Training Set

In this exercise, you will run QikProp in normal processing mode on the imported structures to create an input training file for a QikFit job that will be run in a later exercise.

1. Choose QikProp from the Applications menu.

   The QikProp panel is displayed.

2. Choose Selected entries from the Use structures from option menu.
3. Click Start.
4. The Start dialog box is displayed. In this dialog box you can make settings for the output incorporation and for running the job.
5. From the Incorporate option menu, choose Replace existing entries.
6. In the Name text box, type `ligands`.

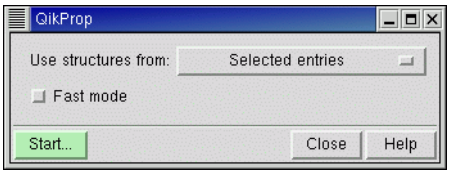

Figure 3.3. The** QikProp **panel.

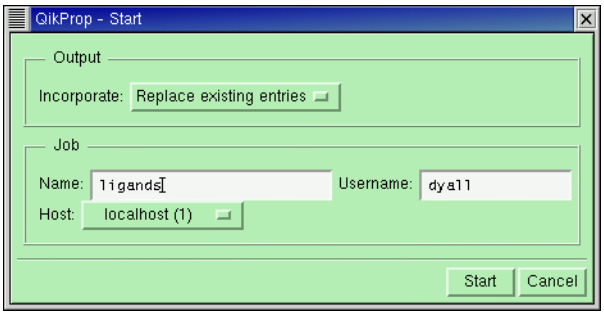

Figure 3.4. The* Start *dialog box.

7. Click Start.

   The Monitor panel is displayed and the job begins running. When the job is complete, the following files should be in your working directory:

```
ligands-out.mae      ligands.qpsa         ligands.mae
ligands.CSV          ligands.out          ligands.log
```

   The file `ligands-out.mae` is a Maestro-formatted file that contains property data and structural information. The `ligands.mae` file is written by QikProp at the start of the job and contains only structural information for the selected entries. The `ligands.out` file lists properties predicted and the limits specified in `QPlimits`. It also includes a section that shows how some of the properties were formed by linear combinations of more basic molecular properties. The `ligands.qpsa` file provides a detailed surface area analysis, and `ligands.CSV` contains all the predicted properties for each structure in comma-separated value (CSV) format.

8. Choose Save As from the Project menu in the main window.

   The Save Project As dialog box is displayed.

9. Name the project `ligs.prj` and click Save.

   The project is saved as `ligs.prj`.

3.4 Running QikProp in Fast Processing Mode

In this exercise, you will run QikProp in fast mode. You will use the same set of structures you processed in normal mode in the previous exercise so that you can compare the output files produced by each mode. To preserve both sets, you will first copy the project.

1. Choose Save As from the Project menu of the main window.

   The Save Project As dialog box is displayed.

2. Enter `comp.prj` under Project and click Save.

   The current project is now named `comp.prj`.

3. Ensure that all entries in the Project Table panel are selected.
4. Open the QikProp panel from the Applications menu in the main window.
5. Choose Selected entries from the Use structures from option menu.
6. Select Fast Mode.
7. Click Start.
8. From the Incorporate option menu, choose Append new entries.
9. In the Name text box, type `QPfast`.
10. Click Start.

    The job runs in a few seconds. When it finishes, the Project Table has 400 entries.

3.5 Sorting Project Table Entries

In this exercise, you will sort the entries in the Project Table to compare the results from fast mode to the results from normal mode. The entries are in groups, so the first step is to ungroup the entries, so that the results for a given molecule are adjacent.

1. In the Project Table panel, select each entry group in turn and choose Ungroup from the Entry menu.
2. Click the Sort button on the toolbar.

   

   The Sort Project Table panel is displayed.

3. Choose Entry Name from the Primary Key list.
4. Choose Ascending from the Order menu.
5. Click Sort All Rows.

   The row for each fast mode entry is now directly underneath the row for the corresponding normal mode entry.

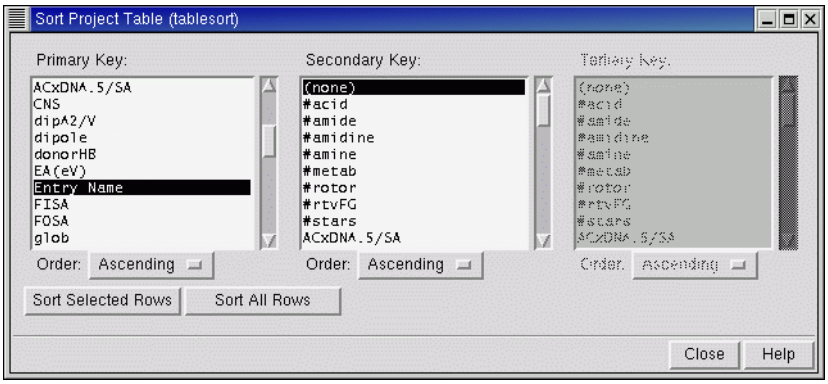

Figure 3.5. The Sort Project Table ***panel.***

6. In the Project Table panel, choose Fit To Header+Data from the Columns button menu on the toolbar.

7. Scroll through the Project Table and compare the fast mode results with the normal mode results.

   Notice that, in fast mode, the dipole moment ("dipole"), dipole^2/volume ("d2ov"), the ionization potential ("ip"), and the electron affinity ("ea") are not predicted. Instead 0.0 is entered for these values. This is because, in the interest of saving time, the PM3 single point calculation is not performed. Notice also that the prediction for the free energy of solvation in octanol ("log poct") differs between fast and normal processing modes because, in fast mode, the QSPR equations are refit without the dipole terms. The `#metab` descriptor also differs between the two modes.

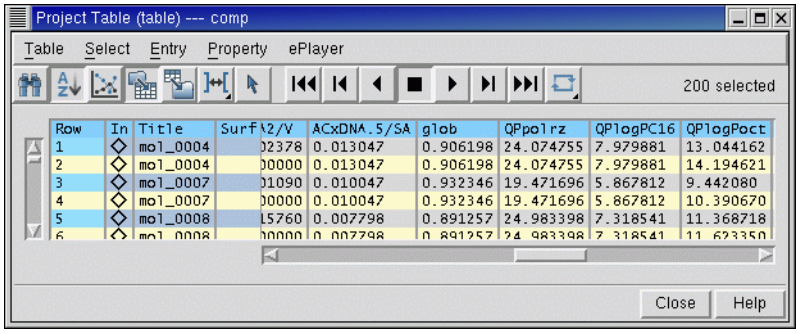

Figure 3.6. The Project Table with fast and normal mode properties sorted.

In later exercises, you will run QikProp in normal mode. To prepare for these exercises:

- In the QikProp panel, deselect Fast Mode.

3.6 Creating a Training File for Use With QikFit

To use QikFit, you must first generate a training file. You can use a spreadsheet program to create a training file, or if you are adding the data as the last column in the CSV file, you can use the Unix utility, `paste`. This exercise uses `paste`. Ordinarily, when creating training files for use with QikFit, you would add a column of new experimental data. For use with this exercise, a list of data has been provided in the `tutorial` directory, in the file `data.txt`. Using the instructions in this exercise, copy the list into the `ligands.CSV` file. In the next exercise, you will use QikFit to fit to this data.

- In your working directory, enter the following command:

  ```
  paste -d , ligands.CSV data.txt > training.CSV
  ```

  A column of data named "exp_data" has been added as the final column, and the resulting expanded file has been saved as `training.CSV`. This new column is our experimental property.

3.7 Using QikProp to Generate a Test File for QikFit

QikFit can calculate the experimental property for molecules in a test file whose name is specified on the command line. At the same time, it can produce a file containing regression coefficients for the experimental value. In this exercise, you will use QikProp to generate the *jobname*`.CSV` file for a molecule called mol_7147, for which there is no "exp_data" value. Instead of running QikProp through the Maestro interface, you will submit the job from the command line. Generally, the command line option is intended for running jobs that consist of large numbers of structures. To create a test file for use with QikFit, enter the following command in your working directory:

```
$SCHRODINGER/qikprop test.mae
```

When the job has finished, the following relevant files should be in your working directory:

```
test-out.mae          test.qpsa             test.out
test.mae              test.CSV              test.log
```

You will use the `test.CSV` file in the QikFit run in the following exercise.

3.8 Running QikFit

In this exercise, you will use QikFit to predict the new "exp_data" property for the test structure mol_7147, whose other properties were predicted in the previous exercise.

To predict the "exp_data" property, enter the following command in your working directory:

```
$SCHRODINGER/qikfit -t test.CSV training.CSV
```

The file `training.CSV` is the file you created in Section 3.6 on page 38. By default, QikFit treats the last column in the training file as the new experimental value. If you want to specify a different column, you can do so with the `-a` option.

When the job is finished, the following files should be in the directory from which you launched QikFit:

```
training.myfits         training.out            training.log
```

The file `training.out` includes the parameter values calculated during each cycle of the regression, statistical data for each cycle, and a list of the experimental and calculated values for each compound. An annotated example of an output file is given in Section 5.4 on page 59. At the end, the new property value is listed:

```
New Predictions

    N    Calc Activity         Compound
      1     1.107678        "mol_7147"
```

3.9 Predicting Properties with the QPmyfits File

While QikFit can predict new properties for a molecule in a test file during a run, it is likely that you will want to use the subsequent *jobname*`.myfits` file to allow QikProp to begin predicting the new property along with the 44 default properties and descriptors for several molecules simultaneously.

The `training.myfits` file that was generated in the last exercise contains the coefficients necessary for QikProp to predict "exp_data" for molecules that are structurally similar to those molecules that are included in the file. In this exercise, you will use `training.myfits` with the same set of training molecules—those from the file `ligs.mae`—that were used in the previous exercises.

1. Rename the `training.myfits` file `QPmyfits`.

   QikProp reads the `QPmyfits` file and uses it for the new property.

2. Choose Close from the Project menu in the main window.

   The previous project is closed and a new scratch project is created.

3. Import `ligs.mae` into Maestro.

   All entries are automatically selected in the Project Table, and the first is included in the Workspace. If they are not selected, choose All from the Select menu in the Project Table.

4. Open the QikProp panel from the Applications menu.
5. Choose Selected entries from the Use structures from option menu.
6. If Fast Mode is selected, deselect it.
7. Click Start.
8. From the Incorporate option menu, choose Replace existing entries.
9. In the Name text box, type `ligands2`.
10. Click Start.

    The Hydrogen/Lone-Pair Fixup dialog box is displayed.

11. Choose Fix For All.

This job produces the following files:

```
ligands2-out.mae        ligands2.out            ligands2.CSV
ligands2.qpsa           ligands2.log
```

This time an additional property appears in the Project Table under the heading "myfit1." Compare the predicted "exp_data" values with the originals in `training.CSV`. Note that the predictions that appear in `ligands2.CSV` are the same as the calculated values reported for each structure in `training.out`.

3.10 Creating a Plot of Project Table Properties

In the next few exercises, you will use the plotting capabilities of the Project Table to examine data produced by QikProp in the previous exercise. In this exercise, you will plot data for a subset of Project Table entries.

1. In the Project Table panel, select a subset of entries to plot by clicking on entry 1, then holding down the Shift key and clicking on entry 20.

   Entries 1 through 20 should now be selected.

2. Click the Plot button on the toolbar.

The Plot XY panel is displayed.

3. Choose New Plot from the Plot menu.

The New Plot dialog box is displayed.

To visualize the range of solvent accessible surface areas (SASAs) in the selected set of entries, you will plot “SASA” versus “Entry Name.”

4. In the Plot Name text box type `SASA`.
5. For the x-axis property, select Entry Name.
6. For the y-axis property, select SASA.
7. Click New.

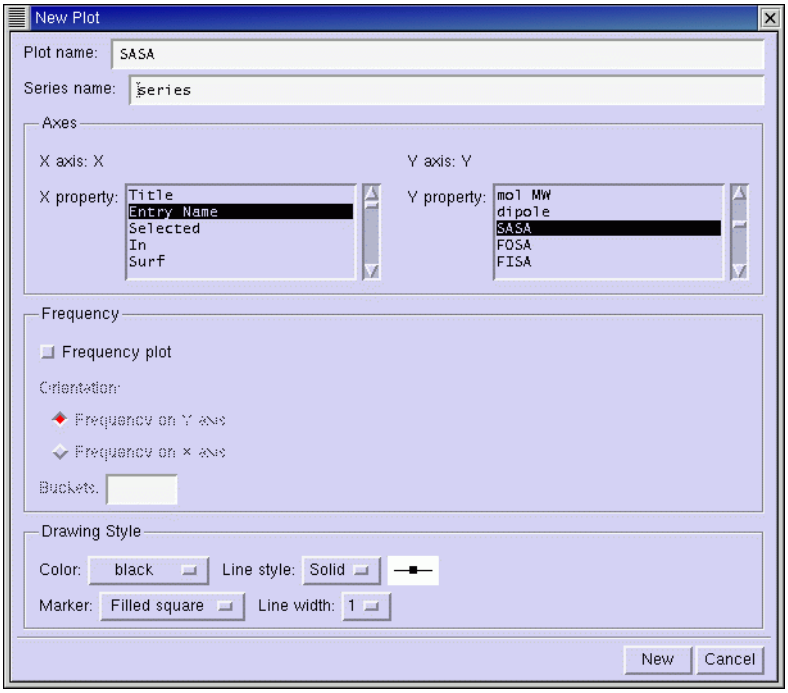

Figure 3.7. The New Plot ***dialog box.***

The plot is displayed in the Plot XY panel. Keep this panel open for use in the following exercise in which you will add a scatter plot.

3.11 Creating a Scatter Plot

In this exercise, you will create a scatter plot of data for all entries in the Project Table, and you will explore some additional features of the Plotting Facility.

1. In the Project Table panel, choose All from the Select menu.
2. In the PlotXY panel, choose New Plot from the Plot menu.

   The New Plot dialog box is displayed. This time you will create a scatter plot to visualize the correlation between two properties.
3. In the Plot Name text box, enter `Scatter`.
4. For the x-axis property, select mol MW.
5. For the y-axis property, select volume.
6. Change the plot style by making selections in the Drawing Style section:
   - For Color choose blue.
   - For Line Style select None.
   - For Marker select Cross.
7. Click New to generate the new plot.

The toolbar contains tools for labeling points, selecting and including in the Workspace the entries that correspond to data points, panning and zooming. Selection and inclusion work exactly as they do in the Project Table. In addition, you can drag over a set of points to select or include them.

The dashed line surrounding a plot indicates that it is selected. You can select plots by clicking on the plot. Once a plot is selected, you can perform various operations on it. To view only the selected plot, choose Display Only Selected Plots from the Plot menu. To view both plots again, from the Plot menu choose Select All, then choose Display Selected Plots.

For more information on plotting, see Chapter 10 of the *Maestro User Manual*.

In the subsequent exercise, you will execute QikSim from the command line, so you may now close Maestro.

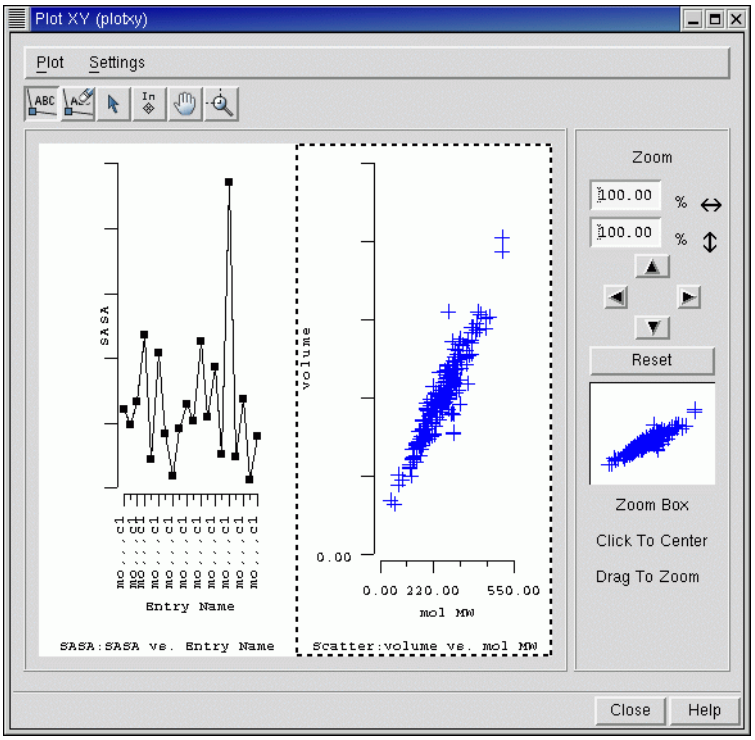

Figure 3.8. The** Plot XY **panel showing two plots.

3.12 Running QikSim and Examining Output Data

QikSim ranks molecules based on structural similarities to a specified reference molecule. As the basis for comparison, QikSim uses the Euclidean distance and the Tanimoto coefficient. To perform a QikSim calculation, you must specify a reference or "probe" structure. The first structure will be used by default, but any structure in the file can be specified with the `-p` option. Use the following instructions to perform the QikSim calculation.

1. Enter the following command:

   ```
   $SCHRODINGER/qiksim -p 10 ligands.CSV
   ```

   The `-p` option selects the tenth structure as the probe molecule. All other structures are compared to structure 10. The job should run in a few seconds.

2. View the contents of `ligands.simout`. An excerpt from this file is shown below:

```
Euclidean Distance                          Tanimoto Coefficient
  N        R^2          Name                  N        TC           Name
   10     0.000 mol_0073                       10     1.000 mol_0073
  130     0.091 mol_1191                      130     0.997 mol_1191
   24     0.292 mol_0159                       18     0.986 mol_0132
   18     0.524 mol_0132                       24     0.983 mol_0159
   11     0.837 mol_0075                       11     0.964 mol_0075
```

The output file should show mol_0073, the tenth structure in the input file, as the first structure for which the Euclidean distance and the Tanimoto coefficient are reported. The Euclidean distance for this molecule should be 0.000 and the Tanimoto coefficient should be 1.000 because the molecule is being compared with itself. The next closest molecule is mol_1191 according to both the Euclidean distance and the Tanimoto coefficient.

Running QikProp

QikProp can be run either from the Maestro GUI or from the command line. Using Maestro to run QikProp enables you to build structures or import structures into Maestro, use them as input for a QikProp job, and then use Maestro's analysis tools and Schrödinger's other computational programs to further investigate lead compounds. You might want to run QikProp from the command line if you are processing a large number of structures, or do not immediately want to use Maestro for analysis of the results.

4.1 QikProp Input

QikProp input structures and files must meet the following criteria:

- Structures must be three-dimensional, not two-dimensional.
- Hydrogen atoms must be explicit. You cannot, therefore, use structures with united-atom types from molecular mechanics programs. If your input contains such structures, Maestro prompts you to correct the hydrogen treatment. There is no check for missing hydrogens when you run QikProp from the command line.
- With the exception of $$$$-delimited MDL SD mol files and Maestro-formatted files, each structure file must contain only one molecule. You can concatenate MDL SD mol (*jobname*`.mol`) files into a single file for input to QikProp. For an example, see the file `$SCHRODINGER/qikprop-v`*version*`/molfiles/many.mol`.
- The allowed elements are H, C, N, O, F, Al, Si, P, S, Cl, Br, and I. In fast processing mode all atom types are permitted, though some results might not be valid if atoms other than the above PM3 atoms are used.
- The number of atoms must not exceed 1000 atoms in the fast mode or 100 non-hydrogen and 150 hydrogen atoms in normal processing mode.

Non-neutral input structures are automatically neutralized prior to processing. In normal mode, if QikProp is unable to neutralize a structure, such as in the case of a protonated quarternary amine, that ligand is skipped and no properties are generated. The output Maestro or SD file displaying calculated QikProp properties includes the original input molecule with properties calculated using the neutralized form.

If your structures do not meet any of the above criteria, you can use LigPrep™ to preprocess the structures so that they are ready for input to QikProp. LigPrep converts structures from 2D to

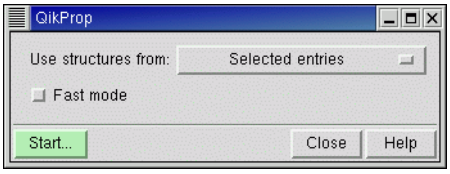

Figure 4.1. The** QikProp **panel.

3D, ensures that explicit hydrogen atoms are used, and performs a molecular mechanics minimization. See the *LigPrep User Manual* for more information on using this product.

QikProp results are not very sensitive to the conformation of the input structure. Where there is variation of the results with conformation, extended structures generally give results that are closer to the experimental data than compact structures. QikProp results are also not very sensitive to the details of structure optimization, though some form of optimized structure should be used for the input.

QikProp processes some additional entries to standard MDL SD mol files, as follows. If there is an entry called `<MOLNAME>`, QikProp reads the following line for the molecule name that is written to all files. If there is an entry called `<MDLNUMBER>` or `<MOLCODE>`, this entry is written after the molecule name in the *jobname*`.out` file.

4.2 Running QikProp From Maestro

To run QikProp from Maestro, launch Maestro according to the instructions in Section 2.2 on page 5, import or build the structures you want to process, correct them if necessary, then select QikProp from the Applications menu. The QikProp panel is displayed.

To launch a job from the QikProp panel, you must specify an input source and a mode (fast or normal). After you make your selections, click the Start button. The Start dialog box is displayed, in which you can specify an incorporating mechanism, a job name, and a host.

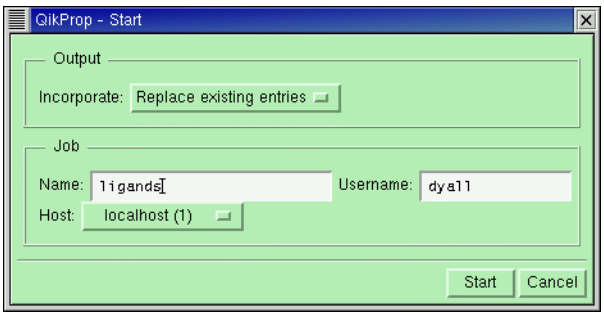

Figure 4.2. The** Start **dialog box.

4.2.1 Setting QikProp Options

Specifying an Input Source

You can use Workspace structures or selected entries in the Project Table panel as input for a QikProp job. The default setting for QikProp job input is Selected Entries. To evaluate structures from a file, you must import the structures into a project using the Import panel, select the desired structures in the Project Table (all new structures are selected by default), and select Selected Entries from the Use structures from option menu in the QikProp panel.

Selecting a Mode

To run in fast mode, select Fast Mode. To run in normal mode, deselect Fast Mode.

4.2.2 Setting Job Options

Selecting a Job Output Incorporation Scheme

QikProp job output can be incorporated into the Project Table in several ways. Select one of the following options from the Incorporate option menu in the Output section.

Replace existing entries

When the job is incorporated, each processed entry is replaced with the QikProp output structure and all of the predicted properties associated with that structure: 40 properties in Fast mode and 44 properties in Normal mode. Generally the output structure does not differ from the structure used as input. However, if QikProp determines that a structure is missing hydrogens, QikProp modifies the input structure and replaces it with the corrected structure.

Append new entries

This is the default option. The output structures are added to the end of the Project Table with their properties. The original entries are not changed.

Do not incorporate

The output files are written to the specified directory, but the new structures and predicted properties do not appear in the Project Table.

Setting the Job Name

To set the job name, enter the name in the Name text box, which is in the Job section. The default name is `qikprop`. Maestro's Job Control Facility does not allow you to run multiple jobs with the same name, so you must select a unique name. To run a job that has the same name as a previous job, you must first remove files associated with the first job.

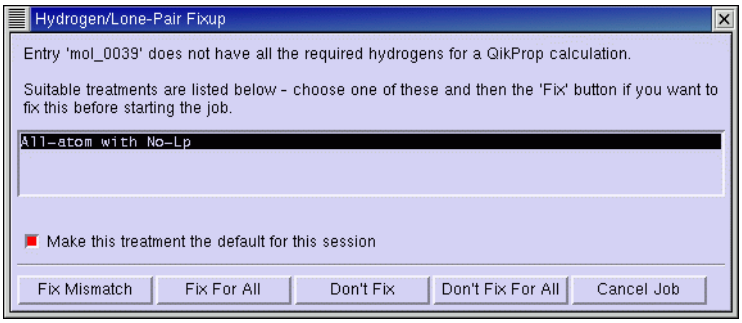

Figure 4.3. The* Hydrogen/Lone-Pair Fixup *dialog box.

Choosing an Execution Host

To run QikProp locally, you do not need to configure your system. Before you can run a QikProp job on a remote host, you must configure the remote host and set up the `schrodinger.hosts` file on the local and remote hosts. Information on these tasks can be found in the *Installation Guide* and the *Job Control Guide*.

To run QikProp on a remote host select a host from the Host option menu. All hosts that have been configured in the `schrodinger.hosts` file appear on this menu.

Specifying a User Name

If your user name on the selected host is different from that on the host on which Maestro is running, you must enter your user name on the selected host in the Username text box.

4.2.3 Fixing Noncompliant Structures

If your input structures are missing hydrogen atoms or other structural information needed by QikProp, the Hydrogen/Lone-Pair Fixup dialog box (Figure 4.3) is displayed when you click Start in the Start dialog box. This dialog box presents a list of valid treatments to fix up the structures, and buttons to select the action to take.

4.3 Running QikProp From the Command Line

To run QikProp from the command line, enter the command:

```
$SCHRODINGER/qikprop [options] input-file
```

The QikProp command line options are described below.

`-fast`	Turn on fast mode.
`-nofast`	Turn off fast mode.
`-nosa`	Do not write the *jobname*`.qpsa` file.
`-n` *lo*`:`*hi*	Process the specified range of molecules from the input file. The lower and upper limits on the molecule number, *lo* and *hi*, must be separated by a colon, with no spaces.
`-HOST` *host*	Run the job on the specified remote host.
`-LOCAL`	Use the current working directory rather than a temporary scratch directory for scratch files.
`-WAIT`	Do not return to the command prompt until the job finishes. The default is to return to the command prompt immediately.
`-NICE`	Run the job at reduced priority.
`-HELP`	Print the acceptable command line arguments and exit.

QikProp input must be a file containing the 3D structure (*x*, *y*, and *z* coordinates and atomic numbers) of one or more molecules, or a file containing a list of input file names. Four file formats are accepted:

- Maestro files (`.mae`)
- MDL mol files, also known as 3D SD files (`.mol`, `.mol2`)
- PDB files (`.pdb`)
- BOSS/MCPRO Z-matrix files (`.z`)

If the file name does not end in one of the suffixes given in parentheses, it is assumed to be a file containing a list of valid input file names. An example of each file type is provided in the directory `$SCHRODINGER/qikprop-v`*version*`/molfiles`. In addition, this subdirectory contains the mol files for 30 common drugs.

4.4 Monitoring QikProp Jobs

QikProp jobs, whether they are run from Maestro or from the command line, are managed by the Schrödinger Job Control Facility. All jobs, whether run from the command line or from Maestro, can be monitored from the Monitor panel, which is shown in Figure 2.1 on page 7.

For information on the Job Control Facility and how to use this facility to monitor QikProp jobs, see the *Job Control Guide*.

When you start a QikProp job from Maestro, the Monitor panel is displayed and the contents of the *jobname*`.out` file are displayed in it as the file is written. The name of the job is placed in the job database portion of the Monitor panel, and the job is automatically selected in the data-

base. By default, Show Jobs From Current Project Only is selected. To show all jobs run from your user name, deselect this option.

When there are no jobs currently running, only the Monitor and Delete buttons are active. While a job is running, the Detach, Pause, Resume, Stop, Kill, and Update buttons are active, but only the Kill button has any effect for QikProp jobs. To terminate a QikProp job, select the desired job in the database list, then click Kill.

4.5 QikProp Output Files

For every completed QikProp job, the following files are created:

jobname`-out.mae`	*jobname*`.CSV`	*jobname*`.mae`
jobname`.qpsa`	*jobname*`.out`	*jobname*`.log`

jobname is the stem of the input file name. If you started the job from Maestro, *jobname* is the name you entered in the Name text box. For example, a job with an input file named `ligs.mae` produces output files named `ligs-out.mae`, `ligs.mae`, and so on. In addition to these standard files, a *jobname*`.warning` file is written if QikProp has difficulty processing a structure.

QikProp places any generated output files in the current working directory. If you run QikProp from Maestro, this directory is the file i/o directory specified in the Directory folder of the Preferences panel, which is shown in Figure 2.6 on page 28.

The default file i/o directory is Maestro's current working directory. To change the file i/o directory preference, open the Preferences panel from the Maestro menu in the main window and select a directory option. The default file i/o directory, as well as most of the other directory choices, is relative. For instance, if you select Maestro's current working directory, Maestro writes output files to the directory from which you launched Maestro, unless you have since entered the `cd` or `chdir` command in the Command Input Area.

The *jobname*`-out.mae` File

Upon completion of a job, QikProp creates the *jobname*`-out.mae` file. This is a Maestro-formatted structure file that contains the QikProp properties generated for each structure as well as structural information. The properties are imported and displayed in the Project Table when the job is incorporated or the file imported. For structures that do not meet the input criteria, no properties are generated in normal processing mode.

The *jobname*`.mae` File

Before Maestro passes a QikProp job on to the Job Control Facility, QikProp first writes a Maestro-formatted *jobname*`.mae` file containing structural information for the molecules to be

evaluated. The structure contained in the *jobname*.mae file depends on the Source of Job Input setting. If you selected Workspace in the QikProp panel, the *jobname*.mae file contains any structures that appeared in the Workspace when the job was started. If you selected Selected Entries, the file contains the structures corresponding to the entries in the Project Table that were selected when the job was started.

The *jobname*.qpsa File

The *jobname*.qpsa file contains additional information for each molecule, including the covalent neighbors list, atom-type assignments, associated atom-type force field parameters, and a solvent accessible surface area (SASA) analysis. At the end of the file, there is a summary giving the number of molecules that were successfully processed and the number of molecules that were not successfully processed, and a report of the processor time used.

The *jobname*.out File

The *jobname*.out file lists the output descriptors, predicted properties, and results of comparisons between the compound's properties and the properties of typical drugs. Each property with a value outside the 95% range for known drugs is marked with an asterisk (*).

The *jobname*.out file also contains a "QP Breakdown" section for each structure. The four quantities listed in this section, log Po/w, -log S, log BB and log PMDCK, are computed as a linear combination of more basic molecular properties. QikProp lists the descriptor values multiplied by their coefficients for each derived property. The predicted property value, listed under Total, is the sum of these. Descriptor values that are above the maximum observed in the training set used to develop QikProp are marked with an angle bracket (<).

The *jobname*.CSV File

The *jobname*.CSV file is a comma-separated-value (CSV) file that lists the 44 QikProp descriptors and predicted properties on one comma-separated line per molecule. The stem of the *jobname*.CSV file corresponds to the job name entered in the Maestro QikProp panel.

The format of the *jobname*.CSV file allows you to import it into spreadsheet programs such as Microsoft Excel, OpenOffice and Gnumeric, or analysis programs such as JMP. You might want to do this to create a training file for use with QikFit. See Section 5.1.1 on page 57 for more information about QikFit training files. When QikProp writes this file, it converts any commas appearing in molecule names to semicolons, and it removes any blanks from molecule names. The *jobname*.CSV file can also be used as input for QikSim jobs.

QikProp makes entries in the *jobname*.CSV file for any molecules with fatal errors. Check the file for these messages after multiple-structure runs.

The *jobname*.warning File

Warnings about potential problems with the input structures are written to the *jobname*.warning file. For instance, warnings are issued when QikProp encounters empty input files, ionic input structures, structures with multiple fragments, and structures with missing hydrogens.

In addition to statements placed in the *jobname*.warning file, if multiple structures are being processed in a single run, any structures with fatal errors (e.g., disallowed atom types) are skipped, and entries are made in the *jobname*.CSV file stating that the processing of the molecules failed. After executing QikProp for multiple structures, you should display the contents of the *jobname*.warning file, if present, and search for any occurrences of empty or failed in the *jobname*.CSV file. If QikProp encounters no problems with the processed structures, a *jobname*.warning file is not generated.

The *jobname*.log File

This file contains a brief summary of the options used to run the job.

4.6 Customizing QikProp with the QPlimits File

The QPlimits file can be used to set the defaults for some QikProp settings or to customize the ranges of values that determine which molecules are flagged as being dissimilar to other known drugs. This file can be copied from the $SCHRODINGER/qikprop-v*version*/data directory to the current working directory or to $HOME/.schrodinger/qikprop.

By default the molecules are flagged as being dissimilar to other known drugs based on comparison to a set of test structures studied by Professor William L. Jorgensen. If you want to use limits other than the default limits, you can edit QPlimits. The current limits for each property are given beneath the name of the property. These limits can be deleted and replaced by user-specified minimum and maximum values. Structures that have properties that do not lie within the limits are flagged with an asterisk in the output files.

You can suppress the *jobname*.qpsa file (called QPSA.out in the example), and choose fast or normal processing modes by changing the settings at the top of the QPlimits file. Changes to the names of the output and CSV files are ignored. Command-line options and Maestro settings supersede any settings made in the QPlimits file. This file is reproduced below.

```
This file lets the user change the names of the output QP.out and QP.CSV files,
suppress the QPSA.out file, choose the fast (no QM) or slower processing modes, and
change the default ranges for 95% of Drugs that are used to determine the "star" count.
Enter name of QP.out file (20 characters max):
QP.out
Enter name of QP.CSV file (20 characters max):
QP.CSV
Suppress output of QPSA.out file? (y or n)
n
Processing Mode (FAST or NORMAL; FAST does not output dipole moment, IP and EA)
normal
Molecular Weight                      standard is:  ( 130.0 / 725.0)
 130.0 / 725.0
Dipole Moment (D)                     standard is:  (   1.0 /  12.5)
   1.0 /  12.5
Total        SASA                     standard is:  ( 300.0 /1000.0)
 300.0 /1000.0
Hydrophobic  SASA                     standard is:  (   0.0 / 750.0)
   0.0 / 750.0
Hydrophilic  SASA                     standard is:  (   7.0 / 330.0)
   7.0 / 330.0
Carbon Pi    SASA                     standard is:  (   0.0 / 400.0)
   0.0 / 400.0
Weakly Polar SASA                     standard is:  (   0.0 / 150.0)
   0.0 / 150.0
Molecular Volume (A^3)                standard is:  ( 500.0 /2000.0)
 500.0 /2000.0
No. of Rotatable Bonds                standard is:  (   0.0 /  15.0)
   0.0 /  15.0
Solute as Donor -    Hydrogen Bonds   standard is:  (   0.0 /   6.0)
   0.0 /   6.0
Solute as Acceptor - Hydrogen Bonds   standard is:  (   2.0 /  20.0)
   2.0 /  20.0
Solute Globularity   (Sphere = 1)     standard is:  (  0.75 /  0.95)
  0.75 /  0.95
Solute Ionization Potential (eV)      standard is:  (   7.9 /  10.5)
   7.9 /  10.5
Solute Electron Affinity    (eV)      standard is:  (  -0.7 /   1.7)
  -0.7 /   1.7
QP Polarizability (Angstroms^3)       standard is:  (  13.0 /  70.0)
  13.0 /  70.0
QP log P  for     hexadecane/gas      standard is:  (   4.0 /  18.0)
   4.0 /  18.0
QP log P  for     octanol/gas         standard is:  (   8.0 /  43.0)
   8.0 /  43.0
QP log P  for     water/gas           standard is:  (   5.0 /  48.0)
   5.0 /  48.0
QP log P  for     octanol/water       standard is:  (  -2.0 /   6.0)
  -2.0 /   6.0
QP log S  for   aqueous solubility    standard is:  (  -6.0 /   0.5)
  -6.0 /   0.5
QP log K hsa Serum Protein Binding    standard is:  (  -1.5 /   1.2)
  -1.5 /   1.2
QP log BB for      brain/blood        standard is:  (  -3.0 /   1.0)
  -3.0 /   1.0
No. of Primary Metabolites            standard is:  (   1.0 /   8.0)
   1.0 /   8.0
```

QikFit

QikFit generates linear regression equations that describe the relationship between a new, user-added experimental property and applicable default QikProp descriptors and properties. Using the following equation, QikFit performs iterative calculation cycles to obtain optimized coefficients for QikProp properties that correlate linearly with the new property:

$$P_j = \sum_i c_i \chi_{ij} + c_0$$

In the above equation, P_j is the property or activity that is to be predicted for molecule *j*, the c_i values are the regression coefficients, χ_{ij} is the *i*th relevant property for molecule *j*, and c_0 is a constant. At the end of each calculation cycle, QikFit discards properties that are not correlated with the new property by more than the significance cutoff value, which is 0.0006 by default.

When the input file contains exactly 45 descriptors and properties (for example *jobname*`.CSV` from QikProp), QikFit generates a *jobname*`.myfits` file. QikProp can then use the values in this file to predict the new property for other similar molecules.

You can use the *jobname*`.myfits` file to integrate coefficients from your own previous QSPR/QSAR analyses into QikProp. This allows you to use those coefficients, rather than the provided default values, to calculate the standard 44 QikProp descriptors and properties.

In addition to the regression analysis described above, QikFit can also perform a trend vector analysis. This is generally used when there are a large number of descriptors and the ratio of molecules to descriptors is not high. For a regression analysis, there should be at least five times as many molecules as descriptors. The trend vector equation has the same form as the regression equations,

$$P_j = \sum_i t_i \chi_{ij}^* + t_0$$

where the t_i are the trend vector coefficients, t_0 is the constant, the χ^* values are the normalized descriptors,

$$\chi_{ij}^* = (\chi_{ij} - \chi_i^{min}) / (\chi_i^{max} - \chi_i^{min})$$

and χ_i^{max} and χ_i^{min} are the maximum and minimum values of property *i*.

5.1 Running QikFit

QikFit cannot be run from Maestro and therefore must be run from a terminal window. Before you run QikFit, ensure that you have set the `SCHRODINGER` environment variable. For more information, see Section 2.2 on page 5. If you want to run a QikFit job on a remote host, you must add the host names that you want to use to the `schrodinger.hosts` file. For more information, see the *Maestro User Manual*.

To run QikFit, enter the command:

```
$SCHRODINGER/qikfit [options] training-file
```

The training file is a comma-separated-value (CSV) file containing the QikProp-calculated properties to be used in the regression, and a column of activities to be fit. See Section 5.1.1 below for more information on training files.

The optional QikFit command line arguments are described in the table below. For more information, see the relevant sections later in this chapter.

`-a` *column*	The column containing the new property data in the training file. A zero value selects the last column. By default, the last column is assumed to hold the new property data.
`-c` *cutoff*	Properties whose contribution to the fit are below the significance cutoff are iteratively removed from the fit. The default cutoff is 0.0006.
`-o` *out-file*	The file in which the output of the analysis is written. The default is *jobname*`.out`.
`-t` *test-file*	An optional comma-separated-value (CSV) file containing the QikProp properties of molecules for which activity values will be predicted using the best-fit coefficients.
`-x` *column*	A column to be excluded from the regression analysis. By default, all columns are used. Only one column to be excluded can be specified. To exclude more than one column, you must edit the file and delete the columns.
`-z`	Perform a trend vector analysis instead of a regression analysis.
`-HOST` *host*	Run the job on the specified remote host.
`-WAIT`	Do not return to the command prompt until the job ends. The default is to return to the command prompt immediately.

5.1.1 The Training File

To use QikFit to predict an additional property, you must have experimental data for enough compounds to comprise a "training set" with which regression equations can be generated. For a regression analysis, it is recommended that you have five times more training compounds than descriptors (approximately 175 compounds). If you have fewer compounds than this, you should perform a trend vector analysis rather than a regression analysis.

The training file must contain molecular properties and descriptors, as well as the new experimental property values for all of the structures.

Requirements for the training file format:

- The training file must be in CSV format. To this file you must add a column containing the new experimental property values.
- The first line must be a header with names for the columns.
- The second and subsequent lines must contain a molecule name or code in the first column, followed by the molecular properties and descriptors.
- The new experimental property should be placed in the last column of each line. If you place it in a different column, you must specify the column number as a command line argument. Use a spreadsheet program to add the data column.
- The number of descriptors that can be used for regression analyses is limited to 250.
- For trend vector analyses, the limit is 2000 descriptors.
- For an analysis of either type, no more than 2500 molecules can be processed in the same file.
- The length of each line is limited to 10,000 characters.

5.1.2 The Test File

You can specify the name of a test file as a command line argument when you run QikFit. This test file should contain descriptors and properties for molecules that are different from, but structurally similar to, those in the training set. If you specify a test file, the optimized regression calculated for the training set is used to predict the new experimental property for the molecules in this file.

Format Restrictions for the Test File:

- The test file must be in CSV format.
- The test file must not have a column containing the new experimental property.

- The test file must have the same number of descriptors and properties as the training file, not including the new experimental property.
- The descriptors and properties must be in the same order as in the training file. This order is the order in which they would appear if the file had been generated using QikProp. (It is possible to generate the test and training files with programs other than QikProp so long as the QikProp property order is preserved.)

5.1.3 Significance Cutoff Values

For the regression analysis only, descriptors with little effect on the fit are pruned in a series of up to 10 calculation cycles. The significance is a measure of the quantitative effect of a change in the descriptor on the sum of squares error (SSE). Selecting higher values for the significance will cause more descriptors to be discarded. The default significance cutoff is 0.0006. If you want to change this value, use the `-c` option to specify an alternative value.

5.2 Output Files

QikFit results are written to the *jobname*`.out` file. If QikFit finds exactly 45 descriptors and properties (44 from QikProp and one experimental value that you added) in the training file, it generates a *jobname*`.myfits` file. This file can be used with QikProp to predict the new property for additional, structurally similar molecules. In addition, a *jobname*`.log` file is produced. This file contains basic information about the job, such as the options that were specified.

The *jobname*`.out` File

The *jobname*`.out` file contains results from the iterative cycles QikFit performs to determine which descriptors in the input file have linear relationships with the new property. The last line of the file contains the predicted new property for the structures in the test file, if one was provided.

Excerpts from a sample output file are provided for both a regression analysis and a trend vector analysis in Section 5.4 on page 59.

The *jobname*`.myfits` File

This file contains coefficients which, when used with QikProp descriptors and properties, can be used to generate the new experimental property for molecules that are structurally similar to those in the training set. A sample of this file is given in Appendix A.

A *jobname*`.myfits` file is only written during a QikFit run when there are exactly 45 descriptors and properties (44 default QikProp descriptors and properties and the additional experimental activity column you have added) in the *jobname*`.CSV` training file.

QikProp does not attempt to determine the identity of the data in the *jobname*.myfits file. It is therefore necessary that the 44 default QikProp descriptors and properties appear in the *jobname*.myfits file, and in the same order in which QikProp generates them.

The *jobname*.myfits file, for each of 4 possible fits, lists the 44 QikProp-predicted descriptors, properties, and coefficients that relate these values to the new experimental property.

5.3 Using QikFit Output to Predict New Properties

You can use the coefficients file *jobname*.myfits with QikProp to predict additional properties, using the following procedure.

1. Open *jobname*.myfits in a text editor.
2. Above the title for the first fit, enter the number of fits you want QikProp to read from the file. For instance, if you have added data for only one new property, enter 1.
3. Below the title for each fit, enter the name of the property you would like to predict.

   This name appears when you print out the file, and it also appears as the column heading in the *jobname*.CSV file.
4. Save the file as QPmyfits into the current working directory.
5. Run QikProp.

See Section 3.9 on page 39 for an example of how to predict properties. See Appendix A for a sample portion of the $SCHRODINGER/qikprop-v*version*/data/myfits.tmp file.

5.4 Sample Regression Output and Commentary

The following is an example of a *jobname*.out file produced by performing a QikFit linear regression for solubility data, log S. The data has been provided in column 2 of the CSV file, logS14.CSV. Commentary is interspersed with the output.

```
QikFit 2.1 - Multiple Linear Regression Analysis

 Incomplete entry - skipping molecule: Ethylfluoride
 Incomplete entry - skipping molecule: norbornane
 Incomplete entry - skipping molecule: norbornene
 Incomplete entry - skipping molecule: tetrachlorodibenzodioxin
 Incomplete entry - skipping molecule: trans-1;2-dichloroethene
 Incomplete entry - skipping molecule: trifluoro-2-chloroethane
```

If an entry in the CSV file is incomplete (missing the activity or the value for a descriptor), or if it contains an illegal character, the entry is ignored and a notice is printed at the top of the file.

```
Column      Contents      Column      Contents
    1       molecule           2       log S
    3       #amine             4       #amidine
    5       #acid              6       #amide
    7       #rotor             8       FOSA
    9       FISA              10       PISA
   11       WPSA              12       volume
   13       donorHB           14       accptHB
   15       ACxDN^.5/SA       16       glob
```

The results for the first cycle of the analysis are now printed using all descriptors except the excluded one, if one was chosen. Since for the example, we chose to exclude the volume, 13 descriptors are used. (The CSV file contains 16 columns, as listed above. Column 1 contains the molecule name and column 2 has the log S values The remaining 14 columns are the descriptors and the volume has been excluded leaving 13 for the analysis.) Standard statistical measures are printed including the r^2 correlation coefficient, rms error, and Fischer F-statistic. A good background reference is given in the Bibliography.

```
Regression Analysis - Cycle  1

          R-Squared         =      0.8233
          RMS Error         =      0.8885
          Average Error     =      0.6765
          Sum of Squares    =    277.0721
          F Statistic       =    121.5341
          Degrees of Freedom=          13

          No. of Data Points=         353
          Average Activity  =     -2.5272
          Exptl Activ Range =    -10.8000 to      2.0600
          Calcd Activ Range =     -9.2915 to      2.2678
          Activity in Column       2
          Signif. Cutoff    =      0.0006

Descriptor           Parameter    Significance   Partial F  Partial R^2
  #amine               0.830900       0.0049       0.0461       0.0001
  #amidine             1.385437       0.0002       0.0005       0.0000
  #acid                0.448397       0.0041       0.8562       0.0024
  #amide               0.836937       0.0035       3.3612       0.0095
  #rotor               0.065276       0.0138       5.0495       0.0142
  FOSA                -0.020641      -0.0079       9.4738       0.0263
  FISA                -0.028545       0.0259      13.1667       0.0362
  PISA                -0.021479       0.0174      65.0079       0.1563
  WPSA                -0.025750       0.0083      69.9055       0.1661
  donorHB              0.556467       0.0528      11.8887       0.0328
  accptHB              0.776197       0.3922       1.1662       0.0033
  ACxDN^.5/SA       -104.176772      -0.0509      10.9283       0.0302
```

```
 glob                -9.314556       0.4962      163.4992       0.3178
 Constant            12.969657       2.7881
```

The next cycle is performed after removal of the descriptors with significance values below the cutoff (0.0006), i.e., #amidine.

```
Regression Analysis - Cycle  2

         R-Squared         =      0.8639
         RMS Error         =      0.7799
         Average Error     =      0.6131
         Sum of Squares    =    213.4976
         F Statistic       =    179.8088
         Degrees of Freedom=          12

         No. of Data Points=         353
         Average Activity  =     -2.5272
         Exptl Activ Range =    -10.8000 to      2.0600
         Calcd Activ Range =     -9.6426 to      1.8926
         Activity in Column       2
         Signif. Cutoff    =      0.0006

Descriptor           Parameter    Significance   Partial F Partial R^2
 #amine               1.167743       0.0013        0.0461       0.0001
 #acid                0.164925       0.0000        0.8562       0.0024
 #amide               0.755102      -0.0003        3.3612       0.0095
 #rotor               0.107415      -0.0055        5.0495       0.0142
 FOSA                -0.022540       0.4668        9.4738       0.0263
 FISA                -0.026933       0.2251       13.1667       0.0362
 PISA                -0.023100       0.3557       65.0079       0.1563
 WPSA                -0.026434       0.0783       69.9055       0.1661
 donorHB              0.828472       0.0173       11.8887       0.0328
 accptHB              0.792982       0.1798        1.1662       0.0033
 ACxDN^.5/SA       -156.746497       0.1268       10.9283       0.0302
 glob                -7.069333       0.7737      163.4992       0.3178
 Constant            11.085043       1.8463
```

The cycles continue until no more descriptors fall below the significance cutoff or until 10 cycles have been performed. The present case terminated on the 3rd cycle and the final results of the analysis are listed. 10 of the original 13 descriptors have been retained:

```
Regression Analysis - Cycle  3

         R-Squared         =      0.8598
         RMS Error         =      0.7915
         Average Error     =      0.6254
         Sum of Squares    =    219.8773
```

```
          F Statistic       =    209.7501
          Degrees of Freedom=          10

          No. of Data Points=         353
          Average Activity  =     -2.5272
          Exptl Activ Range =    -10.8000 to      2.0600
          Calcd Activ Range =     -9.6015 to      1.9240
          Activity in Column       2
          Signif. Cutoff    =      0.0006

 Descriptor           Parameter    Significance   Partial F Partial R^2
  #amine               1.197653       0.0047        0.0461       0.0001
  #rotor               0.132497       0.0048        5.0495       0.0142
  FOSA                -0.023681       0.4508        9.4738       0.0263
  FISA                -0.025538       0.1363       13.1667       0.0362
  PISA                -0.024258       0.3275       65.0079       0.1563
  WPSA                -0.027315       0.0701       69.9055       0.1661
  donorHB              0.821973       0.0372       11.8887       0.0328
  accptHB              0.784791       0.2223        1.1662       0.0033
  ACxDN^.5/SA       -165.056651       0.1082       10.9283       0.0302
  glob                -9.618350       1.2034      163.4992       0.3178
  Constant            13.785027       3.0498
```

The experimental results, the results computed from the final regression equation, and the difference between them are then listed for each molecule.

```
   N    Exptl Activity  Calculated      Delta         Compound
     1    -2.060000     -2.493872      0.433872    (+)-carvone
     2    -2.540000     -1.979838     -0.560162    (E)-2-pentene
     3    -4.000000     -4.082650      0.082650    (R)-limonene
     4     0.620000     -0.046615      0.666615    1-Propanol
     5    -0.070000     -0.421436      0.351436    1-butanol
     6    -2.730000     -2.728530     -0.001470    1-chloropentane
     7    -1.470000     -1.734950      0.264950    1-chloropropane
     8    -1.840000     -1.646043     -0.193957    1-heptanol
   ...
   350    -1.180000     -0.197760     -0.982240    tryptophan
   351    -0.630000     -0.997812      0.367812    valericacid
   352    -1.860000     -2.124633      0.264633    valproicacid
   353    -3.890000     -3.939594      0.049594    warfarin


 QPmyfits file not written - need  36 descriptors.
```

The program states that the QikProp `QPmyfits` file has not been written in this case because the input CSV file did not contain all 44 descriptors and properties. However, it would be easy

to edit a copy of the `myfits.tmp` file to contain the final parameter values from the regression analyses. Renaming the file `QPmyfits` then makes it ready for use with QikProp.

Finally, predictions for any molecules in the test CSV file (`logs14test.CSV`) are listed.

```
New Predictions

N    Calc Activity         Compound
  1    -2.783846        malathion
  2    -0.855569        mannitol
  3    -4.603215        mefenamicacid61-68-7
  4    -2.716188        menthone
  5    -2.456658        meperidine
  6    -2.357517        methylcyclopentane
  7    -3.599530        methylprednisolone
  8    -3.488728        metolazone17560-51-9
  9    -2.123141        metoprolol
 10    -1.292159        metronidazole443-48-1
 ...
```

5.5 Sample Trend Vector Output

The output from trend vector analysis is similar to output from a regression; however, there is only one analysis cycle and all descriptors are used. The following results are for the same data in `logS14.CSV` as above. Again, the volume descriptor has been excluded.

```
QikFit 2.1 - Trend Vector Analysis

 Incomplete entry - skipping molecule: Ethylfluoride
 Incomplete entry - skipping molecule: norbornane
 Incomplete entry - skipping molecule: norbornene
 Incomplete entry - skipping molecule: tetrachlorodibenzodioxin
 Incomplete entry - skipping molecule: trans-1;2-dichloroethene
 Incomplete entry - skipping molecule: trifluoro-2-chloroethane

          R-Squared         =      0.4932
          RMS Error         =      1.5048
          Average Error     =      1.2129
          Sum of Squares    =    794.8508
          F Statistic       =     25.3779
          Degrees of Freedom=          13

          No. of Data Points=         353
          Average Activity  =     -2.5272
          Exptl Activ Range =    -10.8000 to      2.0600
          Calcd Activ Range =     -7.2573 to      1.1300
          Activity in Column       2
```

```
         T Vector Length   =        0.36

 Descriptor           Trend Vec
  #amine               0.005023
  #amidine             0.000134
  #acid                0.018243
  #amide               0.026124
  #rotor              -0.032867
  FOSA                -0.057281
  FISA                 0.081333
  PISA                -0.178242
  WPSA                -0.114821
  donorHB              0.050848
  accptHB              0.016238
  ACxDN^.5/SA          0.040667
  glob                 0.197797
  Constant            -4.440438

    N    Exptl Activity  Calculated      Delta         Compound
      1     -2.060000     -1.667803     -0.392197    (+)-carvone
      2     -2.540000     -1.985081     -0.554919    (E)-2-pentene
      3     -4.000000     -2.528583     -1.471417    (R)-limonene
      4      0.620000     -0.919956      1.539956    1-Propanol
      5     -0.070000     -1.216291      1.146291    1-butanol
      6     -2.730000     -2.319438     -0.410562    1-chloropentane
      7     -1.470000     -1.725160      0.255160    1-chloropropane
      8     -1.840000     -2.143461      0.303461    1-heptanol
      9     -2.380000     -2.444543      0.064543    1-octanol
     10     -0.600000     -1.529228      0.929228    1-pentanol
     11     -2.680000     -2.046645     -0.633355    1-pentene
     ...

New Predictions

    N    Calc Activity         Compound
      1     -3.390014        malathion
      2      0.193206        mannitol
      3     -3.367925        mefenamicacid61-68-7
      4     -1.605173        menthone
      5     -3.590858        meperidine
      6     -1.523822        methylcyclopentane
      7     -1.734992        methylprednisolone
      8     -3.028783        metolazone17560-51-9
      9     -4.302631        metoprolol
     10     -1.189052        metronidazole443-48-1
     ...
```

QikSim

QikSim performs a similarity/diversity analysis for designated ("probe") molecules in a QikProp *jobname*`.CSV` file. During the analysis, QikSim compares the molecule to all the other molecules in that *jobname*`.CSV` file by first computing the Euclidean distance and the Tanimoto coefficient between the probe molecule and each of the other molecules. QikSim then sorts results from most similar to least similar. Such analyses are useful for deriving new leads from given reference compounds.

QikSim is normally run in one of the following two ways:

- A search for molecules that are the most similar to one designated probe molecule (lead). The output gives the ranks of all molecules in the CSV file according to their similarity to the probe molecule. In this manner, QikSim can also be used to find duplicate entries in a database.
- A search for the molecules in the CSV file that are most similar to a set of active compounds. The set must be in consecutive locations in the CSV file. In this case, QikSim computes the averages of the descriptors for the actives and matches the molecules in the CSV file to this average. The output again gives the ranks of all molecules in the CSV file according to their similarity to this probe molecule. In addition, enrichment factors can be computed and written to a designated file. This file includes running tallies of the number of actives that are found in proceeding from the best ranked molecules to the worst. The enrichment *E*(N) for the molecule with rank N is

  $$E(N) = (N_{found}(N)/N_{active}) \times (N_{tot}/N)$$

  where N_{active} is the number of active molecules, N_{tot} is the total number of molecules, $N_{found}(N)$ is the number of active molecules that have been found among the first N molecules. For example, if 50% of the actives are found among the molecules ranked in the top 10%, E(10%) = 0.5 x 10 = 5.

6.1 Command Syntax

To run QikSim, enter the following command from a terminal window:

```
$SCHRODINGER/qiksim [options] CSV-file
```

Here, *CSV-file* is the name of the file to be processed. This file is a comma-separated-values (CSV) file containing the QikProp-calculated properties of the molecules to be compared.

The options for QikSim are described below.

`-b` *molecule*	The molecule number of the first probe molecule in the input CSV file. By default, the first molecule is used.
`-e` *molecule*	The molecule number of the last probe molecule in the input CSV file. By default, this is the same as the first molecule.
`-l` *limits-file*	A file containing the range in which specified property values for each comparison molecule must lie. By default, the `limits.def` file in `$SCHRODINGER/qikprop-v`*version*`/data/` is used.
`-n` *nkeep*	The number of most-similar molecules to report. By default, all are reported.
`-o` *out-file*	The file in which the output of the QikSim analysis is written. The default is *jobname*`.simout`.
`-p` *molecule*	The molecule number of the probe molecule in the input CSV file. The default is `0`, which selects the last row.
`-r [`*enrich-file*`]`	Report enrichment factors for the selected probe molecules, if more than one is selected, and write a CSV file containing enrichment factors. The default file name is *jobname*`-enrich.CSV`
`-HOST` *host*	Run the job on the specified remote host.
`-WAIT`	Do not return the command prompt until the job finishes. The default is to return the command prompt immediately.

6.2 Input

A QikSim analysis requires three input elements:

1. A *jobname*`.CSV` file written by version 2.5 of QikProp. This file must have a complete entry for each molecule. A sample CSV file is provided in the directory `$SCHRODINGER/qikprop-v`*version*`/qiksim/sample`. This file contains QikProp 2.5 results for about 1000 molecules, most of which are known drugs.

   QikSim skips over "failed" entries, but you might want to delete these entries before submission of the file to QikSim.

2. A limits file containing the limits and weights for each QikProp descriptor that is used for the QikSim similarity/diversity analysis. A sample file is provided in the directory `$SCHRODINGER/qikprop-v`*version*`/qiksim/sample`.

   A molecule is output as similar or dissimilar if each of its QikProp descriptors falls between the listed minimum and maximum values. Otherwise the molecule does not

appear in the output. The weights can be used linearly to emphasize or de-emphasize the contribution of a descriptor to the similarity/dissimilarity measure. The file begins with:

```
Descriptor/min, max, weight:
#Stars      Number of stars
    0 1000  0.0
#Amines     Number of non-conjugated amines
    0 1000  0.0
#amidine    Number of amidine and guanidine groups
    0 1000  0.0
#Acid       Number of carboxylic acid groups
    0 1000  4.0
#Amides     Number of non-conjugated amides
    0 1000  0.0
...
```

The Euclidian distance calculation uses the weights from the limits file. The Tanimoto coefficient calculation does not use the values of the weights, but descriptors are ignored if they have zero weight.

The limits file can be specified on the command line. If no file is given, the program looks for a file called `limits` in the current directory. If this file does not exist, the default file, `$SCHRODINGER/qikprop-v`*version*`/data/limits.def`, is used. If you want to use a customized `limits` file, you should copy the default file to your working directory and modify it to suit your purposes.

3. The range of probe molecules in the CSV file and the number of results to report.

   You must look at the CSV file to determine the molecule numbers of the probe molecules. Molecule *n* is in row *n*+1 because the header is in row 1. The default row is the last row.

QikSim skips entries in the CSV file that are incomplete, such as failed entries from QikProp. If the skipped entry is one of the probe molecules, this molecule is deleted from the list of probe molecules. If the list of probe molecules is empty, or if there is only one probe molecule for an enrichment factor calculation, the job stops.

6.3 Output Files

QikSim produces two output files, *jobname*`.out` and *jobname*`-enrich.CSV`.

The *jobname*`.out` file

In the *jobname*`.out` file, the sorted Euclidian distances and Tanimoto coefficients are listed following the input weights from the `limits` file. Below is a sample output file for a set of 25 non-steroidal anti-inflammatory drugs (NSAIDs) from the supplied sample CSV file, which are grouped at the end of the file.

```
**************************************************************
**************************************************************
*************** Similarity/Diversity Analysis ***************
*************** QikSim  Ver. 2.1 - March 2003 ***************
**************************************************************
**************************************************************

Descriptor    Weight
#Stars          0.00
#Amines         2.00
#Amidine        0.00
#Acid           2.00
#Amides         2.00
#Rotor          2.00
Reactv FGs      0.00
CNS Actvty      0.00
MW              1.00
Dipole          1.00
SASA            4.00
FOSA            4.00
FISA            4.00
PISA            4.00
WPSA            1.00
Volume          4.00
HB Donor        5.00
HB Accptr       5.00
Dip^2/V         1.00
ADOSA           0.00
Globularty      4.00
Polariz.        1.00
delG C16        1.00
delG octnl      1.00
delG water      1.00
log P o/w       4.00
log S           1.00
log HERG        0.00
PCaco           2.00
log BB          1.00
PMDCK-Affy      1.00
log Kp          0.00
IP (eV)         0.00
EA (eV)         0.00
#Metabolic      0.00
QPlogKhsa       0.00


Number of Molecules =   908
Number of Actives   =    25


%Actives found in first  5% using Euclidian = 56.00
%Actives found in first  5% using Tanimoto  = 68.00

%Actives found in first 10% using Euclidian = 80.00
%Actives found in first 10% using Tanimoto  = 72.00

       Euclidean Distance                          Tanimoto Coefficient
 N        R^2          Name                    N        TC          Name
 896     0.300 tiaprofenicacid                 896     0.975 tiaprofenicacid
 903     0.434 naprosyn22204-53-1              798     0.970 177785-47-6
 798     0.461 177785-47-6                     891     0.970 caprofen
 894     0.462 indoprofen                      894     0.969 indoprofen
 907     0.516 indomethacin                    903     0.968 naprosyn
 895     0.536 Ketoprofen                      895     0.968 Ketoprofen
 891     0.543 caprofen                        548     0.967 tolfenamicacid
```

```
  306     0.562 Bromazepam                    907     0.964 indomethacin
  801     0.564 185220-03-5                   900     0.963 Naproxen
  898     0.564 fenclofenac                   892     0.962 flurbiprofen
  900     0.599 Naproxen                      899     0.961 fenbufen
  584     0.621 Nevirapine                    898     0.957 fenclofenac
  899     0.654 fenbufen                      845     0.956 mefenamicacid
  845     0.655 mefenamicacid61-68-7          524     0.956 sulindac
  897     0.658 diclofenac                    906     0.955 flufenamicacid
  548     0.665 tolfenamicacid                897     0.955 diclofenac
  892     0.667 flurbiprofen                   93     0.951 bisphenolA
  598     0.684 Loviride(a-APA)               445     0.949 nordazepam
  806     0.693 206662-19-3                   905     0.949 diflumidone
  445     0.709 nordazepam1088-11-5           887     0.949 rofecoxib
  524     0.720 sulindac                      806     0.949 206662-19-3
  410     0.720 Lorazepam                     306     0.948 Bromazepam
  802     0.790 189154-91-4                   802     0.947 189154-91-4
   93     0.807 bisphenolA                    584     0.946 Nevirapine
  460     0.813 PA-824187235-37-6             888     0.946 DFU(COX-2 inhib)
  735     0.823 139339-45-0                   457     0.945 Oxazepam
  734     0.823 139339-45-0                   886     0.945 etoricoxib
  841     0.836 99983-92-3                    270     0.944 A3915
  280     0.838 alprazolam(Xanax)             750     0.944 148472-83-7
  344     0.859 Diethylstilbestrol            801     0.944 185220-03-5
  906     0.863 flufenamicacid                410     0.944 Lorazepam
  457     0.867 Oxazepam                      696     0.943 fluconazole
  753     0.901 153922-98-6                   458     0.943 oxyphenbutazone
  754     0.901 153922-98-6                   790     0.943 172469-92-0
  568     0.905 Young16                       416     0.943 LY153186
  888     0.916 DFU(COX-2 inhib)              588     0.943 MKC-442
...
```

Note that there are some additional NSAIDs such as sulindac that were not included in the grouping of 25 actives, but that do correctly appear near the top of the similarity lists.

The enrichment CSV file

The file `enrich.CSV` has the full enrichment output. It is generated when there is more than one probe molecule and the `-r` option is used. It is a comma-separated value (CSV) file ready for input into a database program. The results can then be displayed to yield plots such as the following plots from JMP.

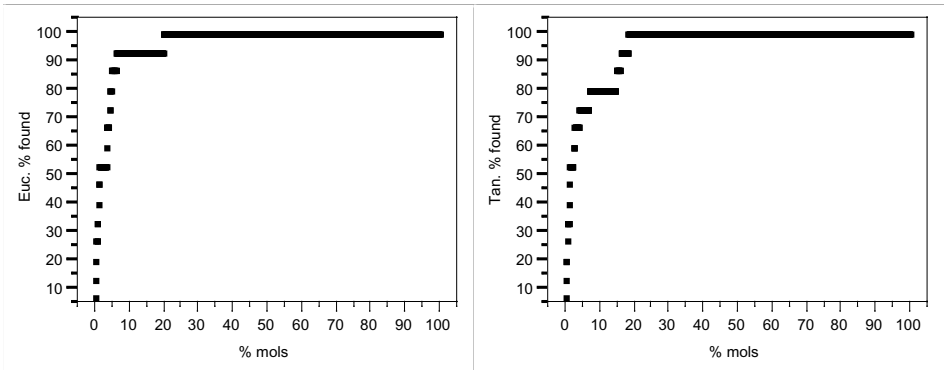

Figure 6.1. Percentage of NSAIDs found vs. percentage of molecules processed.

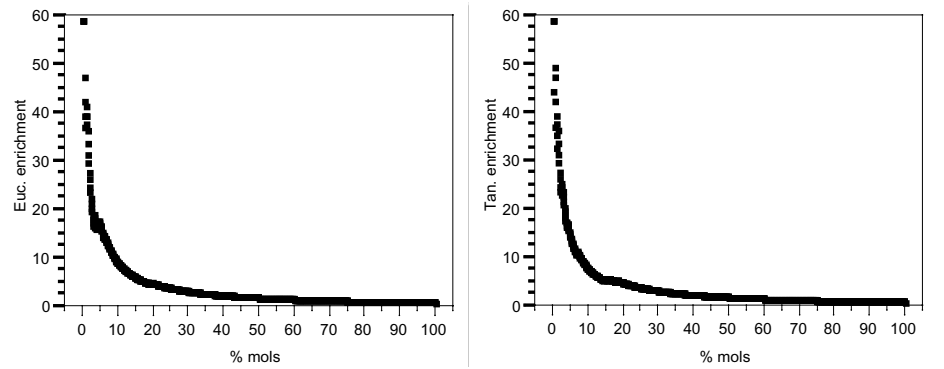

Figure 6.2. Enrichment factors for NSAIDs vs. percentage of molecules processed.

Getting Help

Schrödinger software is distributed with documentation in PDF format. If the documentation is not installed in `$SCHRODINGER/docs` on a computer that you have access to, you should install it or ask your system administrator to install it.

For help installing and setting up licenses for Schrödinger software and installing documentation, see the *Installation Guide*. For information on running jobs, see the *Job Control Guide*.

Maestro has automatic, context-sensitive help (Auto-Help and Balloon Help, or tooltips), and an online help system. To get help, follow the steps below.

- Check the Auto-Help text box, which is located at the foot of the main window. If help is available for the task you are performing, it is automatically displayed there. Auto-Help contains a single line of information. For more detailed information, use the online help.
- If you want information about a GUI element, such as a button or option, there may be Balloon Help for the item. Pause the cursor over the element. If the Balloon Help does not appear, check that Show Balloon Help is selected in the Help menu of the main window. If there is Balloon Help for the element, it appears within a few seconds.
- For information about a panel or the folder that is displayed in a panel, click the Help button in the panel. The Help panel is opened and a relevant help topic is displayed.
- For other information in the online help, open the Help panel and locate the topic by searching or by category. You can open the Help panel by choosing Help from the Help menu on the main menu bar or by pressing CTRL+H.

  To view a list of all available QikProp–related help topics, choose QikProp from the Categories menu of the Categories tab. Double-click a topic title to view the topic.

If you do not find the information you need in the Maestro help system, check the following sources:

- *Maestro User Manual*, for detailed information on using Maestro
- *Maestro Command Reference Manual*, for information on Maestro commands
- *QikProp Technical Notes*, for technical and scientific information on QikProp
- Frequently Asked Questions pages, at https://www.schrodinger.com/QikProp_FAQ.html

The manuals are also available in PDF format from the Schrödinger Support Center. Information on additions and corrections to the manuals is available from this web page.

If you have questions that are not answered from any of the above sources, contact Schrödinger using the information below.

E-mail: help@schrodinger.com
USPS: 101 SW Main Street, Suite 1300, Portland, OR 97204
Phone: (503) 299-1150
Fax: (503) 299-4532
WWW: http://www.schrodinger.com
FTP: ftp://ftp.schrodinger.com

Generally, e-mail correspondence is best because you can send machine output, if necessary. When sending e-mail messages, please include the following information, most of which can be obtained by entering `$SCHRODINGER/machid` at a command prompt:

- All relevant user input and machine output
- QikProp purchaser (company, research institution, or individual)
- Primary QikProp user
- Computer platform type
- Operating system with version number
- QikProp version number
- Maestro version number
- mmshare version number

Bibliography

QikProp

Duffy, E. M.; Jorgensen, W. L. Prediction of Properties from Simulations: Free Energies of Solvation in Hexadecane, Octanol, and Water. *J. Am. Chem. Soc.* **2000,** *122,* 2878-2888.

Jorgensen, W. L.; Duffy, E. M. Prediction of Drug Solubility from Monte Carlo Simulations. *Bioorg. Med. Chem. Lett.* **2000,** *10,* 1155-1158.

Jorgensen, W. L.; Duffy, E. M. Prediction of Drug Solubility from Structure. *Adv. Drug Delivery Rev.* **2002,** *54,* 355-366.

Colmenarejo, G.; Alvarez-Pedraglio, A.; Lavandera, J.-L. Cheminformatic Models To Predict Binding Affinities to Human Serum Albumin. *J. Med. Chem.* **2001,** *44,* 4370-4378.

QikFit

Multiple Linear Regression:

Kleinbaum, D. G.; Kupper, L. L.; Muller, K. E.; Nizam, A. Applied Regression Analysis and Other Multivariable Methods. Third Edition, Duxbury Press: Pacific Grove, 1987.

Miller J. C.; Miller, J. N. Statistics for Analytical Chemistry, 3rd ed.; Ellis Horwood: New York, 1993.

Trend Vector:

Carhart, R. E.; Smith, D. H.; Venkataraghavan, R. Atom pairs as molecular features in structure-activity studies: definition and applications. *J. Chem. Inf. Comput. Sci.* **1985,** *25,* 64-73.

QikSim

Willet, P. Chemical Similarity Searching. *J. Chem. Inf. Comput. Sci.* **1998,** *38,* 983-996.

Sample myfits File

Below is an excerpt from a sample *jobname*`.myfits` file.

```
QikProp 2.5: Use this file for your custom QSAR/QSPR fits to be output by QikProp.
A maximum of 4 fits is allowed. For each fit, enter the coefficients for the 40
descriptors as a decimal number. Fit 1 below is set-up as an example to be the same
as the log Po/w equation used by QikProp for non-hydrocarbons. Activate it by
setting the number of fits to 1.
Enter the number of fits on the next line.
1
****************** Fit 1 *****************
Name for printing - 10 characters maximum
QPlogPo/w
constant
-0.7049
#amine
-0.527
#amidine
0.0
#acid
0.5157
#amide
-0.6256
#rotor
0.0
#rctvFG
0.0
CNS
0.0
MW
0.0
dipole
0.0
SASA
0.0
FOSA
0.0
FISA
-0.006921
PISA
0.001171
WPSA
0.003815
volume
```

```
0.006524
donorHB
-0.2999
accptHB
-0.4870
dip^2/V
0.0
ACxDN^.5/SA
44.36
glob
0.0
QPpolrz
0.0
QPlogPC16
0.0
QPlogPoct
0.0
QPlogPw
0.0
QPlogPo/w
0.0
QPlogS
0.0
CIQPlogS
0.0
QPlogHERG
0.0
QPPCaco
0.0
QPlogBB
0.0
QPPMDCK
0.0
QPlogKp
0.0
IP(eV)
0.0
EA(eV)
0.0
#metabol
0.0
QPlogKhsa
0.0
HumanOralAbs
0.0
QP%HumanOralAbs
0.0
SAfluorine
0.0
```

```
SAamideO
0.0
PSA
0.0
# Ns and Os
0.0
# RuleOf5
0.0
```

Index

120 West 45th Street
32nd Floor
New York, NY 10036

101 SW Main Street
Suite 1300
Portland, OR 97204

3655 Nobel Drive
Suite 430
San Diego, CA 92122

Dynamostraße 13
68165 Mannheim
Germany

QuatroHouse, Frimley Road
Camberley GU16 7ER
United Kingdom

SCHRÖDINGER.